RNA directed epigenetic regulation and modification of the human genome

by

Nicholas Lister

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD) in

Biochemistry and Molecular Genetics

UNSW
SYDNEY

School of Biochemistry and Biomolecular Sciences

Faculty of Science

University of New South Wales

August, 2018

UNSW SYDNEY | Australia's Global University

Thesis/Dissertation Sheet

Surname/Family Name	:	**Lister**
Given Name/s	:	**Nicholas Clive**
Abbreviation for degree as give in the University calendar	:	**PhD**
Faculty	:	**Science**
School	:	**Biotechnology and Biomolecular Sciences**
Thesis Title	:	**RNA directed epigenetic regulation and modification of the human genome**

Abstract 350 words maximum: (PLEASE TYPE)

RNA was originally considered to be a molecule whose primary function was to encode proteins. When it was discovered during the human genome project that the number of protein coding genes was radically lower than previously estimated, the idea was posited that non-coding RNA (ncRNA) could contribute to increasing complexity and account for the observed similarity of number of protein-coding genes between humans and other less complex organisms. Over the last decade, experimental observations have been used to categorise non-coding RNAs by function, which increasingly points towards them being involved in performing a role in regulating the transcriptional and epigenetic states of gene expression.

In this thesis, a number of ncRNA species have been studied to ascertain a greater understanding of the mechanisms involved in transcriptional and epigenetic regulation. I examine short antisense RNA directed transcriptional gene silencing of *Periostin*, resulting in a loss of cellular metastasis. *Periostin* is involved in altering the extracellular matrix to favour increased invasive potential. I investigate the mechanisms involved in the complex PTEN regulatory network, involving RNA:RNA hybrid interactions between an lncRNA pseudogene transcript and a 5'-UTR ncRNA transcript the *PTEN* promoter which regulate gene expression of PTEN via transcriptional and epigenetic regulation. I also studied the deamination DNA repair pathways, resulting in the arisal of cytosine to thymine SNP mutations at loci targeted by shRNAs, antisense lncRNAs, and small antisense RNAs when MBD4 and SMUG repair enzymes have been repressed. Collectively, these data suggest a far more complex role for ncRNAs in the cell has been understood previously, as they appear to play a large part in directing epigenetic regulation and modification of the human genome.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

DECLARATION

‘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 excepted 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 worker at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis in the production of my own work, except to the extent that assistant from others in the project’s design and conception or in style, presentation and linguistic expression is acknowledged.

Signed ……

Date ……

COPYRIGHT STATEMENT

Signed ……

Date ……

AUTHENTICITY STATEMENT

‘I certify that the library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.’

Signed ……

Date ……

LIST OF PUBLICATIONS

Lister, Nicholas C., Matthew Clemson, and Kevin V. Morris. 2015. 'RNA-directed epigenetic silencing of Periostin inhibits cell motility', *Royal Society Open Science*, 2.

Lister, N., G. Shevchenko, J. L. Walshe, J. Groen, P. Johnsson, L. Vidarsdottir, D. Grander, S. F. Ataide, and K. V. Morris. 2017. 'The molecular dynamics of long noncoding RNA control of transcription in PTEN and its pseudogene', *Proc Natl Acad Sci U S A*, 114: 9942-47.

TABLE OF CONTENTS

Page No.

LIST OF TABLES

Page No.

LIST OF FIGURES

Page No.

ACKNOWLEDGMENTS

Firstly, I would like to make a big shout out to Kevin Morris. Thanks for taking me on for both an honours project and PhD way back in the day now. When I rocked up to my interview with you when deciding on a lab to join, I was a little hesitant as you were a bit boisterous and hard to keep up with in my first classes with you. That interview completely changed my mind, as your enthusiasm for both the science you do and your willingness to take me on and show me the ropes in your lab gave me motivation to join your lab. Then when you moved over to the states, you offered me to come with you to continue my experiments, picking me up from the airport and offering your home to me while I found a place to live. I will always appreciate that. It's your dedication to both science and a fun working environment that rubs off on all the people in both of your labs I have had the pleasure of working in. I know our days of foosball are behind us, but I'm always up for a game!

To Paula and Harper, thanks for letting me stay in your home while I sorted out living arrangements. And Fifi, for making sure my bed was comfortable and ensuring there was always a Morris in sight!

Next, I'd like to thank Paul Waters. You were left to take me on as one of your students when Kevin left, and you have been instrumental in helping me put this thesis together. Thanks so much for helping to shape this into the work it has become.

To the original Morris lab crew – Matt, Jess, Caio, David, Albert, Laura, Rosie and Chris, thanks for the fun times we had. It seems like an age ago now, but brings back fond memories, particularly our trip to Lorne. Thanks to Jess in particular for her work on the PTEN paper we published last year, and Matt for showing me the ropes and his work on the SNP paper. Also thanks to Per Johnsson, who has been in touch through email over the last

4 years and has been influential for all of the PTEN work I have done. Your contributions to our PTEN paper provided great insight.

Next, I would like to say thanks to my second COH Morris lab group – Tristan (T-dog), Daniel (Salazar Spinner), Ros (the router?) and Tiina (Madame Pasatiina) - the marooned five; you guys are awesome. I couldn't have asked to meet a better group of friends to work with and explore LA. Also to Juliana (all my visa questions), Valerie (trips to Costco and Mexican food), Surya (Surryeal), Citra (resident dumplings expert), Denis, Ryan, Alicia and Olga, who also made my time in the states special. Thanks to Denis for doing the bioinformatics analysis on our pending SNP paper. Also to Galina, the other Aussie export who has been in both Morris labs, thanks for all your help in and out of the lab. Of course, I can't forget Kevin number 2, my true American housemate. Thanks for giving me a place to stay when I needed it, and for all the good times watching either the CW, Rick and Morty or you playing Zelda. If any of you guys ever come to Australia let me know!

To the Waters lab group – Alex, Will and Kim (who helped me figure out how change word to Australian English thankfully), thanks for your support in the short time I have been in the lab, and the odd foosball game too. Also Shafa, who was there at the start of my PhD.

To my mates outside the lab – in particular Mitchell, Tyson, Todd and Lex, thanks for all the support you guys have given me over the last however many years it's been (10? 15?). Also a big thanks to Jesse, my best mate throughout the course of my 8 years at uni. It's been a long haul, but we are finally ready to graduate! It hasn't been the same without our car trips to and from uni. Always there through thick or thin.

Finally, I'd like to give a huge thanks to my parents, as well as Louise and Jo, who have supported and given me direction through not just my PhD, but my whole life. I wouldn't be where I am today without all your love and support, and I appreciate all you do for me.

ABBREVIATIONS

3’ UTR	3’ untranslated region
5’ Aza-C	5’ azacytadine
5’ UTR	5’ untranslated region
Ago	argonaute
Ago-1	argonaute-1
Ago-2	argonaute-2
APOBEC3A	apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3A
asRNA	antisense RNA
BER	base excision repair
bp	base pairs
C	cytosine
cDNA	complementary DNA
ceRNA	competing endogenous RNA
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation sequencing
circRNA	circular RNA
CO_2	carbon dioxide
CpG	cytosine-phosphate-guanine
CSC	cancer stem cell
DMEM	Dulbecco’s modified Eagle’s medium
DNA	2’deoxyribonucleic acid
DNMT	DNA methyltransferase

DNMT3a	DNA methyltransferase 3a
dsRNA	double stranded RNA
EMT	epithelial-mesenchymal transition
EZH2	enhancer of zeste homologue 2
FACS	Fluorescence-activated cell sorting
FBS	foetal bovine serum
G418	geneticin
H3K27me3	histone 3 lysine 27 trimethylation
H3K9me2	histone 3 lysine 9 dimethylation
HDAC1	histone deacetylase 1
HIV	human immunodeficiency virus
HOTAIR	HOX antisense intergenic RNA
IGV	integrative genomics viewer
IP	immunoprecipitation
kDa	kilodalton
lincRNA	long intergenic RNA
lncRNA	long non-coding RNA
M	molar
MBD4	methyl-CpG binding domain protein 4
miRNA	micro RNA
mRNA	messenger RNA
ncRNA	non-coding RNA
ODN	oligodeoxynucleotide
pAkt	phosphorylation of Akt
PCR	polymerase chain reaction

PI3K	class I phosphatidylinositol 3 kinase
piRNA	piwi-interacting RNA
pg	pseudogene
PIP2	phosphatidylinositol-4,5-biphosphate
PIP3	phosphatidylinositol-3,4,5-triphosphate
POSTN	Periostin
PTGS	post-transcriptional gene silencing
PTEN	phosphatase and tensin homologue
PRC2	polycomb repressive complex 2
qRT-PCR	quantitative real-time PCR
RdRP	RNA-dependent RNA polymerase
RIP	RNA immunoprecipitation
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RNAPII	RNA polymerase II
RNAPIII	RNA polymerase III
RNA-seq	RNA sequencing
rRNA	ribosomal RNA
sasRNA	short antisense RNA
shRNA	short hairpin RNA
siRNA	short interfering RNA
sncRNA	short non-coding RNA
snoRNA	small nucleolar RNA
SNP	single nucleotide polymorphism

SMUG1	single-strand selective monofunctional uracil DNA glycosylase
SOCS	suppressor of cytokine signalling
T	thymine
TBE	tris/borate/EDTA
TDG	thymine DNA glycosylase
Tet	tetracycline
tiRNA	transcription initiating RNA
tRNA	transfer RNA
TGS	transcriptional gene silencing
TSA	trichostatin A
t-UCR	transcribed ultraconserved region
U	uracil
UCSC	university of California, Santa Cruz
UNG	uracil DNA glycosylase
UTR	untranslated region

1 INTRODUCTION

Chapter Contents:

Page No.

List of tables:

Page No.

List of figures:

Page No.

1.1 Scope of Thesis

In this project, I examine a variety of non-coding RNAs (ncRNAs) and their mechanisms involved in both regulating gene expression and modifying the genome content. Paradigm shifts over the last two decades have brought about a newfound excitement for ncRNA based research, as our understanding of an organisms complexity has broadened to include the fraction of the genome that is non-protein-coding. With increasingly advanced genome-wide technologies, we are now able to study and examine functional ncRNAs and the role in epigenetic regulation of the genome, as well as the potential to alter it. ncRNA recruitment of chromatin remodelling complexes are capable of altering the transcriptional activity at target loci. Transcriptional gene silencing is one process studied in depth throughout this thesis, demonstrating the ability to suppress transcriptional activity at genes of interest.

My first study examined if short antisense RNA directed transcriptional gene silencing can affect the phenotype of a gene responsible for increased tumour metastatic and invasive potential. A second investigation aimed to determine the endogenous mechanism involved in the regulation of a tumour suppressor gene, vital for keeping cellular proliferation in check via a complex pseudogene *trans* acting lncRNA network. The final study observes the role ncRNAs play in the deamination DNA repair pathway. The data presents the notion that under certain conditions, such as increased lncRNA expression or impaired deamination repair, that ncRNAs have a role in editing the content of the genome, driving heritable SNP mutations at a particular locus.

My research had two main aims:

1. To investigate the role different ncRNAs play in regulating gene expression through epigenetic mechanisms.
2. To determine if ncRNAs play a role in the emergence of SNPs at target loci.

1.2 Chapter Summary

In **Chapter 1**, I examine the literature surrounding the paradigm shift from protein-coding genes to ncRNAs being responsible for increasing an organisms' complexity. I review how improved genome-wide technologies have led to the discovery of more complex non-coding RNAs, and their roles in epigenetic regulation and transcriptional gene silencing.

In **Chapter 2**, I screen a number of short antisense ncRNAs targeted to *Periostin*, a gene found overexpressed in a number of cancers and in particular metastatic tumours. One novel transcript, named as6, is particularly effective at knocking down *Periostin*, and is found to induce transcriptional gene silencing via transcriptional and epigenetic regulation, resulting in a loss of cellular motility and invasive potential and providing a new potential method for preventing metastasis.

In **Chapter 3**, I disentangle the mechanism behind a long ncRNA (lncRNA) pseudogene regulatory network. This involves an RNA:RNA interaction between an antisense lncRNA pseudogene transcript (acting in trans on a different chromosome), and a 5'-UTR ncRNA transcript that spans part of the *PTEN* promoter. My results indicated this alters the transcriptional and epigenetic regulation of *PTEN*, a tumour suppressor gene, in a manner dependent upon both the structure and sequence of the lncRNA, suggesting a more intricate and complex role for RNA regulation of gene expression than previously thought.

In **Chapter 4**, I tested a variety of different ncRNAs for their ability to direct single nucleotide polymorphisms. shRNAs, antisense lncRNAs, and small antisense RNAs are added to systems where enzymes involved in the deamination repair pathway, MBD4 and SMUG, have been repressed, and are found to induce cytosine to thymine mutations at

particular loci in human cells. This suggests the emergence of SNP mutations in the genome may be occurring in an RNA dependent manner, a novel function not yet ascribed to ncRNAs.

In **Chapter 5**, I summarise the analyses from the previous chapters, and discusses how various ncRNAs play numerous different roles in the cell through their various functions to regulate gene expression in the cell. I provide an overarching model for how ncRNAs are involved in epigenetic regulation and modification of the genome.

1.3 Unravelling the transcriptome

1.3.1 Human genome project

Following the completion of sequencing the human genome in 2001, previous thoughts on the number of protein-coding genes in humans were found to be grossly overestimated (Venter et al. 2001; Lander et al. 2001). Estimates for the number of protein-coding genes typically ranged from 50,000 to 90,000 genes, with wilder estimates even ranging up to 140,000 genes (Roest Crollius et al. 2000). The introduction of new genome-scale technologies that predicted orthologous and paralogous relationships for genes and transcripts, and also between closely related genomes, indicated that the human genome actually encodes for approximately 20,000 protein-coding genes (Goodstadt and Ponting 2006). This number was much lower than previous estimates.

The complexity of an organism was thought to be correlated with the number of protein-coding genes. Non-protein-coding regions were considered “junk DNA”, as they were presumed to not be transcriptionally active, but rather functionally inert (Ohno 1972; Ohno and Yomo 1991). However, this thought process changed, as the genomes of less complex organisms such as mouse (Mouse Genome Sequencing et al. 2002), nematode worm (Goodstadt and Ponting 2006) and chicken (International Chicken Genome Sequencing 2004) were found to have a similar number of protein-coding genes as the human genome. New theories suggested that, instead, the complexity of an organism was controlled by the fraction of the genome non-protein-coding (Taft, Pheasant, and Mattick 2007). A number of these non-coding regions were thought to be transcribed into non-coding RNAs that had function, with transcriptome analyses showing that at least 60-70% of the mammalian genome is transcribed in intricate and complex overlapping patterns (Carninci et al. 2005; Cheng et al. 2005), of which a large portion were not poly-adenylated (Mattick and Makunin 2006; Frith, Pheasant, and Mattick 2005).

1.3.2 The transcriptional landscape

Long thought of as only an informational intermediary between protein and DNA, the multitude of functions that RNA performs within the cell is now being reconsidered as more previously unknown capabilities for RNA are uncovered (Mattick and Makunin 2006). New techniques to identify and understand the role of RNA have resulted in the emergence of new hypotheses of gene regulation. Subsequently, new models in RNA biology have been developed to better understand the increasing complexities of RNA function. New technologies with genome-wide assessment capabilities such as chromatin immunoprecipitation sequencing (ChIP-seq) (Guttman et al. 2009) and RNA sequencing (RNA-seq) (Lister et al. 2008; Mortazavi et al. 2008; Nagalakshmi et al. 2008) have significantly expanded what is known about how gene expression is regulated.

Known cases of infrastructural ncRNA's performing relatively generic functions, such as ribosomal RNAs (rRNAs) (Palade 1955) and transfer RNAs (tRNAs) (Hoagland et al. 1958) involved in protein translation, were documented as early as the 1950's. Then in the 1980's, some RNA were found to be capable of performing enzymatic catalysis (Guerrier-Takada et al. 1983; Kruger et al. 1982), in particular during RNA splicing (Fica et al. 2013). The discovery of more biologically complex and important lncRNAs in the early 1990's, such as H19 (Brannan et al. 1990; Zhang and Tycko 1992) and XIST (Brockdorff et al. 1992; Brown et al. 1992), were considered rare exceptions.

Since the paradigm shift regarding the association between ncRNAs and an organisms complexity in the early to mid-2000's, layers of evidence have highlighted the importance of ncRNA in the regulation of the genome as well as disease states. NcRNAs play important roles in a number of critical functions in the cell, including X-chromosome inactivation (Brown et al. 1992; Brockdorff et al. 1992), epigenetic regulation (Yu et al. 2008), cancer progression (Gupta et al. 2010) and regulation of the cell cycle (Lerner et al. 2009).

1.4 Non-coding RNAs

1.4.1 ncRNAs: roles in the cell

Further advancement in genome-scale technologies over the last two decades have revealed a much richer and more vibrant role for the RNA world in regulating gene expression than previously thought (Carninci et al. 2005; Katayama et al. 2005; Morris and Mattick 2014a). Protein-coding regions represent approximately 1.2% of the euchromatic genome, while at least 60-70% of the genome is transcribed from one or both strands into ncRNAs (Mattick and Makunin 2006). Therefore, transcription is widespread across the genome, with the identification of tens of thousands of ncRNAs (Carninci et al. 2005; Katayama et al. 2005; Guttman et al. 2009; Derrien et al. 2012).

This posits the idea that more non-protein-coding genes exist than protein-coding genes (Yu et al. 2008; Yap et al. 2010). However, most ncRNAs already identified are still yet to be properly studied in an effort to characterise their functional importance and relevance in the cell. NcRNAs are designated into two classes purely based on their length, with short ncRNAs being less than 200bp and lncRNAs being greater than 200bp.

1.4.2 Short non-coding RNAs

Like lncRNAs, there are an abundance of different types of short ncRNAs. The most well studied are microRNAs, which are about 22 nucleotides in length and typically degrade or prevent translation of mRNAs (Fang et al. 2013; Zeng, Yi, and Cullen 2003; Guo et al. 2010). They do this via incomplete target homology, with only 6-8 nucleotides matching the target (Lewis, Burge, and Bartel 2005). Estimates suggest a third of protein-coding genes are regulated by miRNAs (Du and Zamore 2005). Other types of sncRNAs include short interfering RNAs (siRNAs), short antisense RNAs (sasRNAs), transcription initiation RNAs (tiRNAs), piwi-interacting RNAs (piRNAs) and small nucleolar RNAs (snoRNAs).

Table 1-1 SncRNA subclasses

SncRNA subclass	Size (nucleotides)	Main functions
siRNA (short interfering RNA)	~22	RNAi (Waterhouse, Graham, and Wang 1998)
Transcription initiating RNA (tiRNA)	18	Nucleosome positioning; epigenetic regulation of transcription (Taft et al. 2011)
piRNA (Piwi-interacting RNA)	26-30	Epigenetic and post-transcriptional silencing of transposons (Girard et al. 2006)
snoRNA (small nucleolar RNA)	60-300	2'-*O*-ribose-methylation; Pseudouridylation of target RNAs (Kiss 2001)
miRNA (microRNA)	~22	Translational Suppression (Lee, Feinbaum, and Ambros 1993)
sasRNA (short antisense RNA)	~22	Transcriptional gene silencing (Ackley et al. 2013)

SiRNAs (and shRNAs – short hairpin RNAs) can perform the process of RNAi (RNA interference) in cells. This causes gene silencing following the introduction (transfection or transduction) of sense-antisense pairs (Waterhouse, Graham, and Wang 1998). The enzymes Drosha and Exportin 5 cleave and export double-stranded RNA (dsRNA) precursors out of the nucleus (Lee et al. 2003). There, Dicer processes them into smaller dsRNA moieties (21–24 nt). One strand gets loaded into the Ago component of the RNA-induced silencing complex (RISC) (Bernstein et al. 2001). This complex is directed to target loci due to complete sequence complementarity with the loaded RNA, which then undergoes RNA destabilization via Ago-2 (Argonaute-2) mediated cleavage (Ramachandran and Ignacimuthu 2013; Turner and Morris 2010). This post-transcriptional process is thought to occur in the cytoplasm.

1.4.3 Long non-coding RNAs

Since the 1990's, the idea of functional lncRNAs with many different roles has existed. Novel research into this previously unknown field resulted in the discovery of what were at the time considered rare cases of functional lncRNAs. The first of these was H19 in 1990, which is monoallelically expressed from the maternally inherited allele as a result of genomic imprinting (Brannan et al. 1990; Zhang and Tycko 1992). This was followed by the discovery of XIST in 1991, which mediates X-chromosome inactivation (Brown et al. 1991; Brockdorff et al. 1992).

These first lncRNA discoveries were mostly found by large scale sequencing and screening of full-length cDNA libraries (Tsai et al. 2010; Gibb, Brown, and Lam 2011). LncRNA discovery was made easier after the introduction of microarray technologies, such as the HOX tiling array. On this platform the HOTAIR (HOX antisense intergenic RNA) lncRNA was identified, a regulator of HoxD via PRC2 (polycomb repressive complex 2) (Gupta et al. 2010). As technologies have improved, in particular high-throughput RNA

sequencing, so has the number of lncRNAs discovered increased (Carninci et al. 2005; Katayama et al. 2005).

Numerous classes of lncRNAs exist, including pseudogenes, antisense RNAs (asRNAs), long intergenic RNAs (lincRNAs – which fall between protein coding genes), long intronic RNAs (which are within the introns of another transcript) and transcribed ultraconserved regions (T-UCRs - which are observed throughout the genome and are entirely conserved between human, rat and mouse genomes).

circRNA (circular RNA) biology is a relatively recent field that has seen an explosion of publications in the last 5 years (Kristensen et al. 2018). These are non-linear ncRNAs transcribed from thousands of genes. First discovered in the late 1970's, they were thought to be the result of splicing errors (Hsu and Coca-Prados 1979; Cocquerelle et al. 1993). circRNAs are highly conserved and form stable covalently closed RNA circles, with the potential to regulate gene expression (Kristensen et al. 2018).

Table 1-2 LncRNA subclasses

LncRNA subclass	Example
Pseudogene	PTEN (Poliseno et al. 2010)
asRNA (antisense RNA)	TSIX (Lee, Davidow, and Warshawsky 1999)
lincRNA (long intergenic RNA)	HOTAIR (Gupta et al. 2010)
Long intronic RNA	SAP18 (Louro et al. 2008)
T-UCR (transcribed ultraconserved region)	uc.33 (Rahman et al. 2004)
CircRNA (circular RNA)	ciRS-7 (Memczak et al. 2013)

Most lncRNA's are polyadenylated, though a growing number of non-polyadenylated lncRNA transcripts have been identified (Gibb, Brown, and Lam 2011; Gutschner and Diederichs 2012). LncRNAs have been found to have a number of roles in several biological processes. It was found that antisense lncRNA's are responsible for the epigenetic regulation of viral transcription in HIV (Saayman et al. 2014b). The main functions discussed here are epigenetic regulation of transcription, but other functions include stem cell pluripotency, cell cycle, mRNA splicing and metabolism (Gibb, Brown, and Lam 2011; Prensner and Chinnaiyan 2011; Rinn and Chang 2012). LncRNAs can be either *cis* or *trans* acting, and alter gene expression at transcription or post-transcriptional levels (Prensner and Chinnaiyan 2011; Guil and Esteller 2012). The mechanisms employed include interacting with transcription factors and chromatin remodelling proteins, as well as regulating mRNA decay (Rinn and Chang 2012; Tsai et al. 2010; Khalil et al. 2009).

1.4.3.1 Antisense RNAs

asRNAs express from the reverse strand to sense transcripts, so have complementary sequences to the sense transcript (except in the case of bidirectional promoters). They can overlap with 5' and 3' UTRs, introns and exons, and gene promoters. Some of the most well studied asRNAs include TSIX (Lee, Davidow, and Warshawsky 1999), which is antisense to XIST at the X-inactivation centre, and HOTAIR (Gupta et al. 2010). Advances in genome-wide technologies have demonstrated that antisense transcription occurs throughout the genome (Katayama et al. 2005). Estimates for protein-coding gene with an associated antisense transcript range from 20-40% (Chen et al. 2004; Engstrom et al. 2006).

Most examples of asRNAs are known to act *in cis* (Yu et al. 2008; Morris et al. 2008; Faghihi et al. 2010), though an increasing number of *trans* acting asRNAs are being discovered (Rinn et al. 2007; Hawkins and Morris 2010a). Mechanisms for asRNA gene regulation include altered mRNA splicing (Beltran et al. 2008), imprinting (Lee, Davidow,

and Warshawsky 1999) and epigenetic modulation (Yu et al. 2008). They also play a role in disease, as the tumour suppressor gene p53 is known to undergo asRNA regulation (Mahmoudi et al. 2009). Characterising asRNA mediated regulation currently requires investigation of specific candidates to determine possible function.

1.4.3.2 Pseudogenes

Pseudogenes are duplicated relics of genes that no longer have protein coding potential. Loss of protein-coding potential occurs via mutations that cause either premature stop codons or a frame shift. The first pseudogene discovered was in the 1970's, with a truncated 5S DNA gene (coding for 5S rRNA) found with homology to the parent gene (Jacq, Miller, and Brownlee 1977). Previously, pseudogenes were thought to be non-functional, though this hypothesis changed at the start of the 2010's (Poliseno et al. 2010; Hawkins and Morris 2010a). They are known to have a role in regulating parent gene expression, such as the PTEN pseudogene (Johnsson, Ackley, et al. 2013b; Poliseno et al. 2010; Hawkins and Morris 2010a). Pseudogenes are classified based on how they originated from their parent gene.

Table 1-3 Pseudogene subclasses

Pseudogene Subclass	How it is derived
Processed	mRNA that has been spliced, reverse transcribed and inserted into genome at new loci
Unprocessed	Gene duplication event
Unitary	Previously active gene now inactive through mutation

The number of pseudogenes in the genome is similar to the number of human protein-coding genes (14,694 based on human GENCODE version 27, GRCh38). Over millions of years of primate evolution, expressed pseudogenes are conserved (Khachane and Harrison 2009). Among GENCODE annotated pseudogenes, processed pseudogenes outnumber unprocessed pseudogenes by almost three times (10,704 vs 3,469 respectively) (human GENCODE version 27GRCh38), thought to be due to an eruption of retrotranspositional activity in primates around forty million years ago. This corresponds with Alu expansion in primates (Ohshima et al. 2003).

Unitary pseudogenes are least numerous, around seventeen times less than unprocessed pseudogenes (206) (human GENCODE version 27, GRCh38). Pseudogenes can make molecular analyses particularly complicated, as there is high sequence similarities between the parent gene and the pseudogene. As such, sequence similarities must be taken into consideration when performing analyses where pseudogenes are present.

1.5 Transcriptional gene silencing

Transcriptional gene silencing (TGS) is a mechanism first observed in tobacco plants which had been transformed with two vectors, each containing genes conferring either kanamycin or hygromycin resistance, but did not exhibit the transgene phenotype of the first vector once the second vector was transformed into the cells (Matzke et al. 1989; Wassenegger et al. 1994). Small double-stranded RNA's targeted to the transgene were observed to induce epigenetic changes, such as DNA methylation, which led to transgene suppression (Matzke et al. 1989; Napoli, Lemieux, and Jorgensen 1990). This became known as transcriptional gene silencing, and can produce stable, long-term epigenetic modifications that are inherited by daughter cells, termed transgenerational epigenetic inheritance (Morris 2009b, 2015).

For TGS of a specific gene to occur, an ncRNA must have a target locus it shares sequence complementarity with. LncRNA endogenous regulation of a loci can be usurped and effectively bypassed via the transfection/transduction of an exogenous sncRNA to perform TGS, such as a sasRNAs (short antisense RNAs), siRNAs or shRNAs (Turner and Morris 2010). sasRNAs and shRNAs can be synthesised *in vitro* and ligated into the pCDNA3-U6M2 plasmid, which can be expressed from the RNA polymerase III promoter (U6) when transfected/transduced. Once expressed, these small RNAs direct epigenetic complexes to the selected site of interest (Morris 2009c; Han, Kim, and Morris 2007). Further studies showed that the sasRNA's could target the promoter (or upstream of), or even overlapping the 5' UTR region of the transcript (Yu et al. 2008; Morris 2009b).

Plants use processed siRNAs transcribed from RNA Polymerase V to direct both DNA methylation and TGS (Matzke and Mosher 2014). RNA-dependent RNA polymerase (RdRP) activity is required to boost this RNA directed TGS, whereas humans do not have this polymerase, allowing for *cis* and *trans* acting RNA directed TGS (Matzke and Mosher 2014). This is similar to *S. cerevisiae*, which also lack RdRP activity (Camblong et al. 2007) and uses antisense transcripts to direct TGS (Morris et al. 2008).

The asRNA is thought to be attracted to target by sequence complementarity to the target locus promoter. The resultant DNA:RNA hybrid structure recruits the Ago-1 (Argonaute-1) protein at targeted promoter sites in the first 24 hours in siRNA and sasRNA mediated TGS (Turner and Morris 2010). In *S. Pombe*, the sole Ago protein is known to guide TGS by recruiting methyltransferases to its target (Matzke and Birchler 2005; Lippman and Martienssen 2004). In mammalian cells, its role is less clear (Peters and Meister 2007). The importance of Ago-1 is highlighted by it knockdown via RNA interference (RNAi), which results in the target gene functioning normally, as opposed to TGS occurring (Turner and Morris 2010; Younger and Corey 2011). DNA methylation by TGS is not as well

understood in humans as it is in plants, as methylation does not always occur in human TGS. This may be due to the length of RNA targeting at the promoter, effectiveness of RNA delivery, or the presence of other RNAs or proteins at the promoter (Weinberg and Morris 2016).

The chromatin around the TGS target site can be characterized using ChIP. At 48-72 hours post-transfection there is significant increases in histone 3 lysine 27 trimethylation (H3K27me3) and histone 3 lysine 9 dimethylation (H3K9me2) levels, both indicative of epigenetically silenced loci (Weinberg et al. 2006; Morris 2009c; Turner and Morris 2010). Accordingly, the protein EZH2 (enhancer of zeste homologue 2) is enriched at target loci. EZH2 is a histone methyltransferase that adds a methyl group to lysine residues in histone tails, such as at H3K27. DNA methylation transpires after three to four consecutive days of asRNA targeting (Morris 2009c).

The outcome of these epigenetic modifications is a folded heterochromatin structure that is inaccessible to RNA polymerase II (RNAPII) (Yu et al. 2008) and transcriptionally silent. Another protein found at TGS target sites includes histone deacetylase 1 (HDAC1), which removes acetyl groups from histones that inhibit methylation at lysine residues. Also observed at target sites is DNA methyltransferase 3a (DNMT3a), which adds methyl groups to DNA at cytosine-phosphate-guanine (CpG) residues (Turner and Morris 2010; Han, Kim, and Morris 2007).

DNMT3a is considered a *de novo* DNA methyltransferase (DNMT), whereas another member of the family, DNMT1, is involved in maintenance and can recognise hemimethylated segments of newly formed DNA (Di Ruscio et al. 2013). RNAs are known to bind and interact with these DNMTs, such as *ecCEBPA*, a lncRNA which binds to DNMT1 and prevents methylation of the *CEBPA* gene locus (Di Ruscio et al. 2013). DNMT3a, HDAC1 and EZH2 are known to form an epigenetic regulatory complex in humans (Fuks et

al. 2001; Turner and Morris 2010). Many of the same chromatin modifying proteins are seen between sasRNA/siRNA TGS and lncRNA regulation, as seen in **Figure 1-1** below.

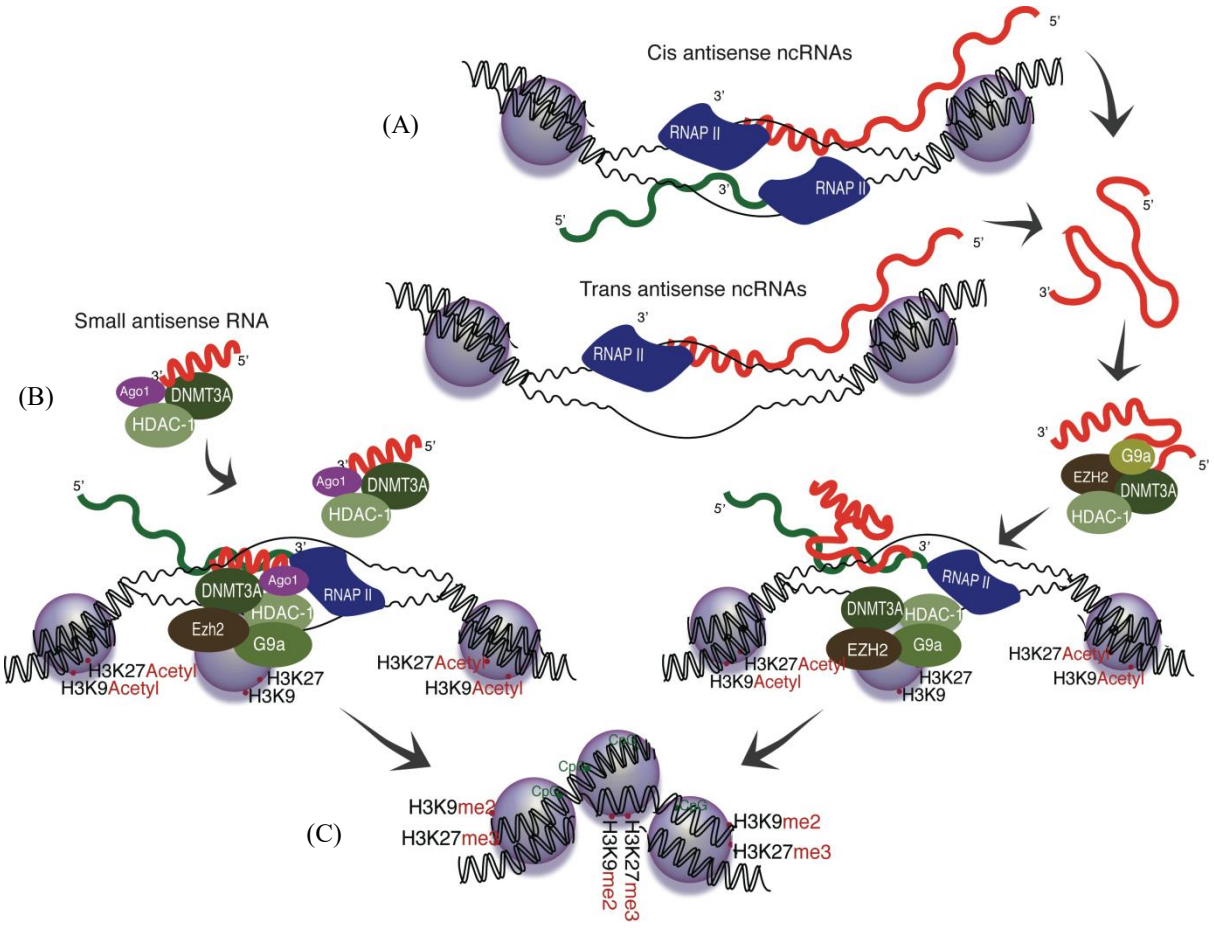


Figure 1-1 Transcriptional Gene Silencing

Schematic depicting endogenous *cis* and *trans* acting antisense ncRNA transcriptional modulation on the right (**A**), and sasRNA directed TGS on the left (**B**). These ncRNA's interact with a number of proteins in complex including Ago1, DNMT3a, HDAC1 and EZH2 to epigenetically remodel target loci via DNA methylation and chromatin modification, resulting in a heterochromatic structure (**C**) and transcriptional silencing. Figure adapted from (Lister, Clemson, and Morris 2015).

1.6 Therapeutic benefits of TGS

RNA directed regulation of gene expression has great therapeutic potential. To be able to target a specific transcript involved in a disease state of interest and stably silence the expression of that gene long-term presents an opportunity to modify the function of the cell towards a disease-free state. This methodology of transcriptional regulation could also be applied to a gene that has been inactivated through deregulation caused by a disease state (e.g. tumour suppressor genes). These genes could be reactivated by silencing their endogenous regulatory ncRNAs (Morris 2009c).

Therapeutic advantages of TGS are that it requires targeting for only a few days and has the potential to result in stable epigenetic change (and thus modified gene expression) that can then be inherited by daughter cells. One particular advantage of small single-stranded RNA targeting is specificity, since only the site of interest will be targeted. Non-specific drugs can produce unwanted side effects (Morris 2009c; Wei, Yang, et al. 2009; Wei, Wei, et al. 2009; Turner and Morris 2010). These single-stranded RNAs can be modified, as alterations to their phosphate sugar backbone can result in greater stability that is less likely to degrade, permitting long term stable targeting (Morris 2009c).

The delivery mechanism and stability of the cell remain as hindrances currently preventing the widespread use of RNA based TGS. Alterations to the backbone of the single-stranded RNA can improve resistance to degradation, though they are still highly unstable in serum, as they undergo hydrolysis and attack by nucleases. Also, modified RNAs also don't seem capable of inducing TGS. The modifications can block transcription, but cannot induce the epigenetic changes associated with TGS, suggested unmodified RNAs are required (Knowling et al. 2012).

Lipid and polymer-based nanoparticles delivery mechanisms struggle to deliver the RNA to the target cells without toxic side effects. Viral delivery mechanisms also result in

permanent viral DNA integration into the host genome (Turner and Morris 2010). Further research into increasing the stability of single-stranded RNAs and alternate mechanisms for delivery is necessary before RNA induced TGS can become a viable therapeutic prospect. However, resolving these issues could lead to the extensive use of TGS in treatment of various disease states. Progress has already begun, as siRNA induced TGS has been shown to work *in vivo* via lentiviral transduction, capable of inhibiting HIV-1 replication in mice (Suzuki et al. 2013).

1.7 Periostin

1.7.1 Characterisation of Periostin Protein

Periostin is a protein secreted into the extracellular matrix. It is also known as a matricellular protein (an extracellular matrix molecule which can moderate and alter interactions between the extracellular matrix and the surrounding cells). Matricellular proteins demonstrate high expression levels during tissue development, either prenatally or during tissue repair and regeneration, or in the presence of a disease state (Hamilton 2008). The most well-known isoform, variant 1, is 835 amino acids long and approximately 90 kDa. Five other variants are known, which have alternative splicing in the C-terminus of the protein. Only three of these variants have been fully sequenced (Morra and Moch 2011; Morra et al. 2011). *Periostin* is located on human chromosome 13q (chr13:38,132,190-38,159,363 on UCSC Genome Browser v290).

Periostin contains a fourfold repeating domain with sequence homology to fasciclin 1 (a protein found in insects). It also has a highly-conserved N-terminus signal peptide, a cysteine rich domain, and a hydrophilic C-terminus that is variable depending on isoform (Takeshita et al. 1993; Morra and Moch 2011). It is part of the fasciclin family of proteins. When the whole C-terminus is removed, motility and cell migration are prevented (Ruan,

Bao, and Ouyang 2009). The C-terminus interacts with other extracellular matrix proteins, including collagen I/V, fibronectin and heparin (Sun, Zhao, et al. 2011). The fasciclin-like domains act at the cell membrane, where they interact with integrins. This interaction is attributed to functions including cell motility and cell adhesion (Morra and Moch 2011; Kanno et al. 2008).

The binding of Periostin to various integrins activates the AKT/PKB and FAK-mediated signalling pathway. This promotes cell survival and angiogenesis, and is also attributed to tumorigenesis (Tischler et al. 2010; Morra and Moch 2011; Gritsenko, Ilina, and Friedl 2012). *Periostin* is normally expressed during tooth and bone maintenance, heart development, and in embryonic tissues such as the periosteum and periodontal ligaments. It is also overexpressed during tissue repair (Sun, Zhao, et al. 2011; Tischler et al. 2010; Hamilton 2008).

1.7.2 Periostin Expression in Cancer

Periostin has been observed to be overexpressed in many types of cancer, due to tumorigenic events (Kudo et al. 2007; Tischler et al. 2010). It is reported to be secreted by stromal fibroblasts in tumours (Malanchi et al. 2012; Kikuchi et al. 2008). *Periostin* is easily detected in breast and ovarian cancer when compared to normal breast and ovarian tissue, where it cannot be detected (Choi et al. 2011; He et al. 2011; Kharaishvili et al. 2011; Tian et al. 2012; Zhang et al. 2010). *Periostin* levels are also increased in cases of primary breast cancer metastasising to bone (Contie et al. 2011; Kharaishvili et al. 2011; Kyutoku et al. 2011; Sasaki et al. 2003; He et al. 2011).

In pancreatic cancer, the stromal cells in the pancreas are responsible for secretion of Periostin, as opposed to the tumour cells themselves. This may be a contributing factor in providing a microenvironment suited to cancer progression (Ruan, Bao, and Ouyang 2009; Baril et al. 2007; Ben et al. 2011; Duner et al. 2010; Streit et al. 2009). In 80% of colon cancer

patients, *Periostin* is overexpressed. Two-thirds of these patients have a five-fold increase in *Periostin* levels, and the other third a 10-fold increase (Bao et al. 2004; Ben et al. 2009; Li et al. 2007). All reported cases of liver metastases demonstrated *Periostin* overexpression in the metastatic tumour (Li, Wang, and Chi 2013; Wu, Wang, and Zhang 2013).

Periostin has also been reported in some lung cancers, with expression levels correlating with tumour size and stage, as well as invasion levels (Ruan, Bao, and Ouyang 2009). In the absence of metastasis, no *Periostin* expression is observed in the lung; however, when a tumour has metastasised to the lung from elsewhere in the body, *Periostin* is observed in the lung stroma. *Periostin* expression correlates with metastatic efficiency (Malanchi et al. 2012; Takanami, Abiko, and Koizumi 2008). Non-small cell lung cancer (NSCLC) patients have a poorer prognosis if *Periostin* presents in their tumours. This suggests a role for *Periostin* in tumour progression and angiogenesis (Soltermann 2012; Morra et al. 2012; Soltermann et al. 2008).

In prostate cancer, *Periostin* epithelial expression correlated with the tumour grade and stage, whereas stromal *Periostin* expression was only correlated to the tumour grade (Tischler et al. 2010; Nuzzo et al. 2012; Sun, Song, et al. 2011; Sun, Zhao, et al. 2011; Tsunoda et al. 2009; Tian, Bova, and Zhang 2011). *Periostin* overexpression has also been reported in thyroid cancer, neuroblastoma and squamous cell carcinoma (Ruan, Bao, and Ouyang 2009; Kudo et al. 2007; Tischler et al. 2010; Gunia et al. 2013; Kudo et al. 2012).

1.7.3 Mechanisms for Enhanced Tumour Survival

Tumorigenesis is an intricate process, involving numerous genetic and epigenetic alterations. *Periostin* expression is integral for tumours to obtain many characteristics that permit cancer cell invasion and survival. These include 1) the ability to avoid apoptosis. This is done by promoting survival under the stress conditions that present themselves in a tumour microenvironment (e.g. insufficient nutrients and hypoxia); 2) inducing angiogenesis (via

integrins and the FAK-mediated signalling pathway), providing nutrients to the tumour; 3) proliferation independent of growth factor signalling. This is done by promoting re-entry into the cell cycle (via integrins and the AKT/PKB pathway); and 4) becoming resistant to anti-proliferative signals (Ruan, Bao, and Ouyang 2009; Kudo et al. 2007; Forsti et al. 2007; Gritsenko, Ilina, and Friedl 2012). The interaction between Periostin and integrins regulates genes that enable survival in tumour cells (Ruan, Bao, and Ouyang 2009; Kudo et al. 2007). *Periostin* is also overexpressed under hypoxic conditions, resulting in cell survival via the AKT/PKB pathway (Liu and Liu 2011; Ouyang et al. 2009; Erkan et al. 2009).

Compared to normal tissue, during tumorigenesis the extracellular matrix is altered in the microenvironment surrounding the tumour. Therefore, Periostin may play a role in the modulation and reorganisation of the extracellular matrix structure to augment tumour invasion and metastatic potential (Morra and Moch 2011; Hamilton 2008; Tumbarello, Temple, and Brenton 2012). Epithelial-mesenchymal transition (EMT) is a process that happens during tumour metastasis as well as embryonic development. If *Periostin* is overexpressed during EMT, epithelial cancer cells gain the ability to invade, differentiate and metastasise (Kudo et al. 2007; Sun, Song, et al. 2011; Puglisi et al. 2008; Soltermann 2012; Siriwardena et al. 2006).

A cell line derived from a colon cancer known to rarely metastasise, was modified to overexpress *Periostin* and implanted in immunocompromised mice. The tumours that developed presented increased metastasis to other organs *in vivo*. This demonstrates the heightened metastatic potential of tumours during *Periostin* overexpression (Bao et al. 2004). Periostin protein has been observed via chromatin immunoprecipitation (ChIP) analysis and mass spectrometry to interact with other extracellular matrix molecules, including collagen I/V and fibronectin (via integrins), suggesting that Periostin has a role in fibrillogenesis (the

formation and regulation of collagen) (Hamilton 2008). Fibrillogenesis provides the ability for a cell to control connective tissue structure and function in both healthy and disease states.

In summary, Periostin appears critical for improved tumour cell survival, as well as metastatic and proliferative ability. There is an important correlation between high *Periostin* expression and poor cancer patient prognosis, demonstrating its potential as a therapeutic gene silencing target so as to prevent tumour metastasis.

1.7.4 Relationship Between Periostin and Cancer Stem Cells

Cancer stem cells (CSCs) are a group of cells derived from cancer that are responsible for long term tumour survival, and possibly remission (Malanchi et al. 2012). It is still not clear whether CSCs derive from a committed progenitor cell or from normal stem cells that have undergone mutation (Malanchi and Huelsken 2009). Transplanted CSCs can cause complete tumour regeneration. Both CSCs and overexpression of *Periostin* result in increased metastatic potential and tumour cell invasion (Malanchi and Huelsken 2009; Malanchi et al. 2012; Wang and Ouyang 2012; Xu et al. 2012).

Mortality from cancer most commonly occurs due to tumour metastasis (Michaylira et al. 2010). New microenvironments previously unexposed to tumours will not provide favourable conditions for secondary tumour invasion. This is because a new microenvironment will have an intact extracellular matrix and will only present a small quantity of CSCs (Malanchi et al. 2012; Malanchi and Huelsken 2009; Wang and Ouyang 2012). Metastasis is not an efficient process and will result in the death of most of the cancer cells during invasion (Wang and Ouyang 2012).

1.8 PTEN

1.8.1 Characterisation of PTEN

The phosphatase and tensin homologue (*PTEN*) is a tumour suppressor gene located on chromosome 10q23 (Steck et al. 1997). It is often mutated (Ali, Schriml, and Dean 1999; Saal et al. 2008), deleted (Ali, Schriml, and Dean 1999) or transcriptionally silenced by epigenetic mechanisms (Garcia et al. 2004; Mirmohammadsadegh et al. 2006) in cancer. As such, it is one of the most frequently down-regulated tumour suppressor genes in cancer. Downregulation of *PTEN* gene dose has been reported to increase the severity and incidence of epithelial cancers (Alimonti et al. 2010; Trotman et al. 2003). The *PTEN* gene has been reported to be mutated in glioblastoma (Walker et al. 2004). Susceptibility to cancer was shown to increase even when minor changes in *PTEN* expression are observed (Alimonti et al. 2010). In cases of metastasis and more advanced tumours, *PTEN* expression is generally completely absent (Ali, Schriml, and Dean 1999).

PTEN is a known negative regulator of the oncogenic PI3K/Akt pathway (Xing 2010). *PTEN* is an antagonist of PI3K (class I phosphatidylinositol 3-kinase) (Carracedo and Pandolfi 2008). PI3K acts as a catalyst for PIP2 (phosphatidylinositol-4,5-biphosphate) phosphorylation into PIP3 (phosphatidylinositol-3,4,5-triphosphate). *PTEN* dephosphorylates PIP3 into the inactive PIP2 (Carracedo and Pandolfi 2008). When *PTEN* is silent during the development of cancer, PIP3 accumulates and hyper-activates the PI3K/Akt pathway (Poliseno and Pandolfi 2015). This triggers pAkt (phosphorylation of Akt), which initiates the downstream cascade signals that upregulate proliferative and cell survival capabilities (Carracedo and Pandolfi 2008).

Stable expression of *PTEN* is important in preventing tumour development. As such, *PTEN* expression is highly regulated in the cell at the transcriptional, post-transcriptional and translational levels. *PTEN* was found to work with *p53*, also a tumour suppressor gene. p53

binds to the *PTEN* promoter to initiate transcription (Stambolic et al. 2001). If PTEN is completely absent during early tumour development, p53 is known to cause tumour cell senescence (Chen et al. 2005). pAKT phosphorylation of the MDM2 protein leads to degradation of p53 in the cell nucleus (Mayo and Donner 2001), negating its tumour suppressor role. *PTEN* can also be translationally suppressed by miRNAs (Alimonti et al. 2010), such as miR21, which is induced in a number of tumour types, causing destabilization of *PTEN* mRNA (Meng et al. 2007).

1.8.2 PTENpg1 regulation of PTEN

MiRNA sponging was first reported endogenously in *PTEN* (Poliseno et al. 2010). CeRNA (competing endogenous RNA) networks between miRNAs, *PTEN* and its pseudogene (*PTENpg1*), work to post-transcriptionally regulate *PTEN* expression (Poliseno and Pandolfi 2015). *PTENpg1* lncRNA, located on chromosome 9 (Teng et al. 1997), acts as a miRNA sponge by sequestering *PTEN* targeting miRNAs away from *PTEN* mRNA. Overexpression of the 3' UTR of *PTENpg1* lncRNA was shown to increase *PTEN* mRNA stability and, subsequently, levels of PTEN protein. In contrast, suppression of the 3' UTR of *PTENpg1* led to *PTEN* mRNA destabilization (Poliseno et al. 2010).

Functional asRNAs were reported to originate from the *PTENpg1* promoter locus, which regulate *PTEN* expression both transcriptionally and post-transcriptionally (Johnsson, Ackley, et al. 2013b, 2013a). It was found that the *PTENpg1* promoter is bidirectional, giving rise to two different isoforms of this antisense transcript, called α and β. An RNA:RNA interaction between the *PTENpg1* sense and *PTENpg1* asRNA β antisense transcripts were able to alter the physical location of the *PTENpg1* sense transcripts in the cell, which in turn effects miRNA sequestration and subsequently levels of PTEN protein (Johnsson, Ackley, et al. 2013b, 2013a). The *PTENpg1* asRNA α acts *in trans*, localising to the *PTEN* promoter

and attracting the epigenetic remodelling proteins DNMT3a and EZH2, resulting in chromatin compaction and TGS of *PTEN* (Johnsson, Ackley, et al. 2013b, 2013a).

1.9 Deamination Repair Pathway

NcRNAs are capable of directing somatically heritable epigenetic changes that can be passed on to daughter cells (Hawkins et al. 2009a). As such, ncRNAs may be in part responsible for mediating transgenerational epigenetic inheritance if this were to occur in the germ line, upon which selection processes could then act (Morris 2015). NcRNAs recruit epigenetic remodelling proteins EZH2 and DNMT3a to target loci, causing DNA methylation of cytosine residues (Hawkins et al. 2009a; Johnsson, Ackley, et al. 2013b; Weinberg et al. 2006).

Methylated cytosines usually undergo deamination in the cell, a process whereby APOBEC3A (apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3A) protein deaminates methylated cytosine to a uracil residue (Wijesinghe and Bhagwat 2012; Carpenter et al. 2012; Suspene et al. 2013), which then undergoes mismatch repair by TDG (thymine DNA glycosylase) (Hashimoto, Hong, et al. 2012). If this process is disrupted, some methylated cytosine residues will be hydrolysed, producing ammonia and resulting in a uracil residue replacing the original cytosine residue in the DNA (Morris 2015; Neddermann et al. 1996). After DNA replication, this uracil will be recognised as a thymine residue, and will become fixed in one of the two daughter strands, resulting in a heritable SNP (Lutsenko and Bhagwat 1999; Shen, Rideout, and Jones 1994).

There are several other glycosylase proteins that also play a role in the deamination repair pathway. These include MBD4 (methyl-CpG binding domain protein 4) (Hashimoto, Liu, et al. 2012; Kondo et al. 2005), SMUG (single-strand selective monofunctional uracil DNA glycosylase) (Kemmerich et al. 2012) and UNG (uracil DNA glycosylase) (Hashimoto,

Hong, et al. 2012), each of which repair cytosine residues that have been deaminated to uracil by APOBEC3A (Budworth and McMurray 2013). HDAC-1, involved in ncRNA epigenetic regulation, was reported to interact directly with TDG, suggesting a link between ncRNA directed epigenetic silencing and the deamination repair pathway (Turner and Morris 2010; Morris 2009c).

Previously considered a random event, cytosine to thymine residue deamination occurs twice as fast at methylated cytosine residues than non-methylated cytosine residues (Shen, Rideout, and Jones 1994). This suggests that the deamination process has an inherent bias toward cytosine methylation at CpG regions, with cytosine to thymine residue changes being the most common single nucleotide mutation reported (Neddermann et al. 1996; Alexandrov et al. 2013).

1.10 Summary

Over the past two decades, our understanding of the RNA regulatory world has completely changed. Paradigm shifts have led us away from thinking of much of the genome as "junk information", and towards the realisation that the RNA holds the key to increased complexity within an organism. Advancing technologies have allowed us an inside look at the diverse and intricate regulatory methods devised by the cell to regulate epigenetic processes, chromosomal organisation, transcription and translation via ncRNAs. However, there is still much to uncover, with tens of thousands of ncRNAs with unassigned functions all taking part in the fascinating transcriptional landscape of the cell. Disentangling this vast RNA network is an enormous task. This thesis aims to contribute to this immense undertaking, in an effort to improve our understanding of ncRNA epigenetic regulation and modification of the genome.

2 RNA targeted inhibition of metastasis

Chapter contents:

Page No.

List of tables:

Page No.

List of figures:

Page No.

2.1 RNA-directed epigenetic silencing of Periostin inhibits cell motility

Authors:

Nicholas C. Lister, Matthew Clemson and Kevin V. Morris

Current status of paper:

Published

Date paper accepted for publication or anticipated date of publication:

14 May 2015

Name of Journal:

Royal Society Open Science 2015; 2(6):140545

Extent to which research is your own:

I performed all the experiments for this manuscript. These include examining the expression profile of *Periostin*, a nuclear run-on assay, ChIPs for DNMT3a and RNAPII, drug treatments, as well as a cell count viability assay and a scratch assay. I performed all statistical tests and data analysis for these experiments.

Your contribution to writing the paper:

I wrote the manuscripts first draft, as well as generating all of the data figures (figure 1b, d, all of figures 2 and 3).

Comments:

The following manuscript has been appropriated from a published paper. This manuscript describes a novel RNA molecule targeted to the promoter of a gene, *Periostin*. When overexpressed, *Periostin* is involved in increased cell motility and metastasis in numerous cancer types, including breast, ovarian, lung, colon, head and neck, pancreatic, prostate, neuroblastoma and thyroid cancers. The novel RNA molecule found here is observed to induce transcriptional gene silencing via epigenetic regulators such as deacetylases and methyltransferases, resulting in a loss of cellular motility in human cancer cells.

2.2 Manuscript sent for publication

RNA-directed epigenetic silencing of Periostin inhibits cell metastasis

Nicholas C. Lister[1], Matthew Clemson[1], and Kevin V Morris[1,2,*]

[1]School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, Australia

[2]Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA, United States, 92037

*Correspondence author, kmorris@unsw.edu.au

Keywords: Periostin, metastasis, RNA, Transcription, gene silencing, Epigenetic

2.2.1 Abstract

The over-expression of Periostin, a member of the fasciclin family of proteins, has been reported in a number of cancers and in particular in metastatic tumours. These include breast, ovarian, lung, colon, head and neck, pancreatic, prostate, neuroblastoma and thyroid cancers. It is thought that Periostin plays a major role in the development of metastases due to its apparent involvement in restructuring of the extracellular matrix to create a microenvironment favouring invasion and metastases, angiogenesis, independent proliferation, avoidance of apoptosis and the ability for cells to re-enter the cell cycle. As such we reasoned that targeted suppression of Periostin at the promoter and epigenetic level could result in the stable inhibition of metastases. We found here that promoter directed small antisense non-coding RNAs (sasRNAs) can induce transcriptional gene silencing (TGS) of Periostin that results ultimately in a loss of cellular metastasis. The observations presented here suggest that metastasis can be controlled by transcriptional and epigenetic regulation of Periostin offering a potentially new and novel manner to control the spread of cancerous cells.

2.2.2 Introduction

Periostin is an approximately 90kDa matricellular protein capable of altering interactions between the extracellular matrix and surrounding cells (Morra and Moch 2011; Morra et al. 2011). Its structure contains a fourfold repeating domain that shows sequence homology with fasciclin 1, a protein found in insects, which interact with integrins at the plasma membrane to accommodate motility and cell adhesion isoform (Takeshita et al. 1993; Morra and Moch 2011; Kanno et al. 2008). Binding of Periostin to various integrins activates the AKT/PKB and FAK-mediated signalling pathway, which promotes angiogenesis and cell survival, and is also involved in tumorigenesis by enabling enhanced tumour cell survival and proliferative and metastatic ability (Tischler et al. 2010; Morra and Moch 2011; Gritsenko, Ilina, and Friedl 2012; Kudo et al. 2007). There is a significant correlation between high *Periostin* expression and poor prognosis in cancer patients (Ben et al. 2009; Choi et al. 2011; Forsti et al. 2007; Nuzzo et al. 2012; Tsunoda et al. 2009), indicating its therapeutic potential as a gene-silencing target in an effort to reduce metastatic tumour formation.

One pathway for inducing stable long-term transcriptional gene silencing (TGS) involves targeting a genes promoter with small antisense non-coding RNAs (sasRNAs) (Morris 2009c) (**Figure 2-1a**). This process involves the sasRNAs recruitment of epigenetic modifying complexes to the RNA targeted locus, ultimately rendering the targeted promoter loci inaccessible to RNA Polymerase II (RNAPII) (Turner and Morris 2010; Yu et al. 2008; Han, Kim, and Morris 2007). One advantage to sasRNA directed TGS is that it can produce lasting, stable epigenetic modifications that are inherited by daughter cells (Morris 2009b). As such we hypothesised that sasRNAs targeted to direct TGS of *Periostin* could inhibit invasive and metastatic potential of cancer cells. We report here a sasRNA that can direct TGS of *Periostin* resulting in the inhibition of cell metastasis. The sasRNA reported here

may prove therapeutically relevant as a means for stably silencing *Periostin* activity and metastasis of tumour cells.

2.2.3 Results

2.2.3.1 Periostin promoter targeting with sasRNAs

Altered expression of *Periostin* is reported in a number of different forms of cancer (Ruan, Bao, and Ouyang 2009; Tischler et al. 2010; Kudo et al. 2007). Endogenous expression of *Periostin* was tested in four different cell lines. Results from qRT-PCR indicated that PC3 cells from a prostate cancer cell line had the highest expression of Periostin mRNA when compared to the other cell lines assessed (**Figure 2-1b**). Expression was approximately ten times greater in PC3 cells than that observed in HEK293 cells, which demonstrated the next highest expression. By comparison, CFPAC and HeLa cells showed negligible expression of *Periostin*. Next, we sought to determine the susceptibility of *Periostin* to promoter targeted sasRNAs. To determine the ability to direct TGS of *Periostin*, five sasRNA sequences directed to the *Periostin* promoter were designed (**Figure 2-1c**), synthesised (Ackley et al. 2013) and screened in PC3 cells. All five sasRNAs were capable of suppressing *Periostin*, with one candidate, as6, demonstrating a robust and significant repression of *Periostin* expression (**Figure 2-1d**). Collectively, these data suggest that *Periostin* is susceptible to sasRNA-directed TGS, similar to previous observations with other genes (Ackley et al. 2013; Hawkins et al. 2009a; Janowski et al. 2005; Janowski et al. 2006; Kim et al. 2008; Kim et al. 2007; Lim et al. 2008; Morris et al. 2004; Roberts et al. 2012; Singh et al. 2014; Suzuki et al. 2011; Suzuki et al. 2008; Turner et al. 2012).

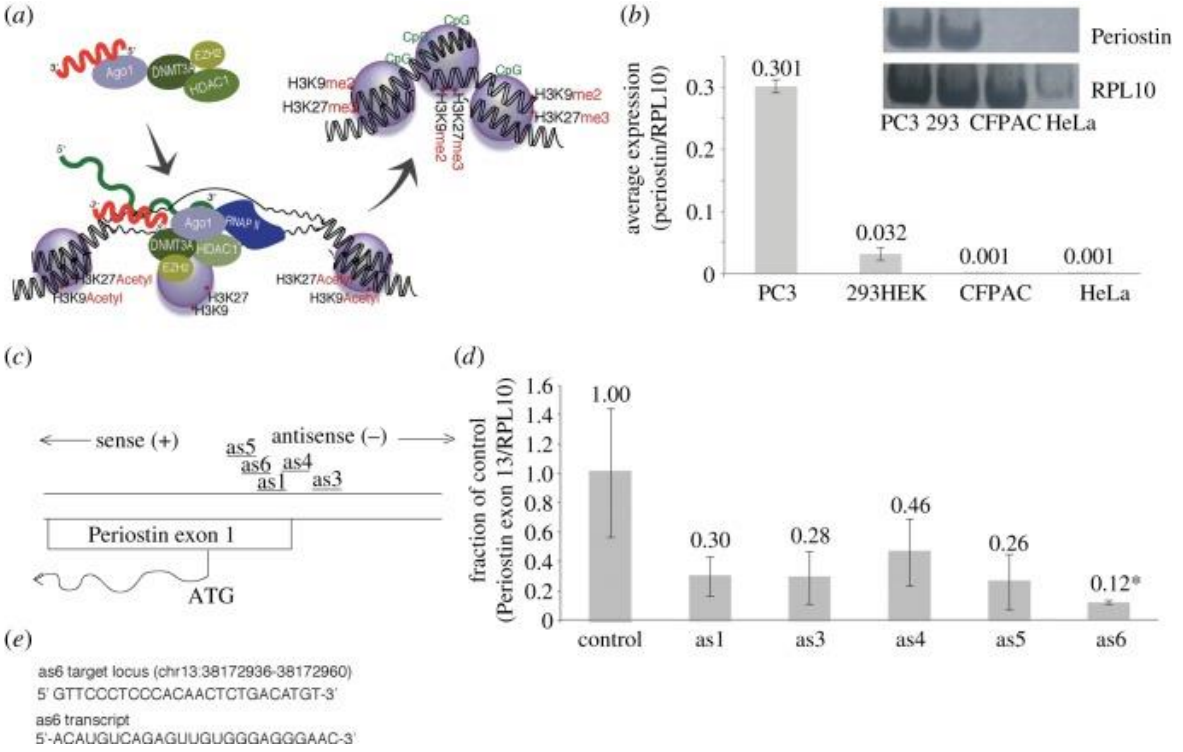


Figure 2-1 Characterisation of *Periostin* Expression and Knockdown

(**a**) Schematic depicting sasRNA directed TGS. The sasRNAs interact with Argonaute 1 (AGO1), DNMT3a, Histone Deacetylase 1 (HDAC1) and Enhancer of Zeste 2 (EZH2) to epigenetically remodel target loci resulting in chromatin compaction and transcriptional silencing. (**b**) qRT-PCR assay of endogenous expression of *Periostin* transcripts in various cell lines, normalised to RPL10 (n=3). Inset is a polymacrylamide gel of qRT-PCR products from the various cells. (**c**) Schematic is shown depicting the sasRNA target loci in the *Periostin* promoter. (**d**) *Periostin* expression in sasRNA transfected PC3 cells 72 hours post-transfection. (**e**) The sequences for the target locus and as6 transcript. The average of triplicate treated cultures are shown with the standard errors of the mean and p values from a paired T-test. Calculations are relative to parent U6M2 plasmid. n.s. indicates data is not significant, * $p < 0.05$ by two-tailed t-test. Error bars indicate s.e.m.

Periostin knockdown is a result of sasRNA directed TGS

The target sequence for as6 in the *Periostin* promoter overlaps the 5' UTR of *Periostin* (**Figure 2-1e**) and the observed suppression of *Periostin* by as6 may be post-transcriptional in nature. In order to confirm the observed reduction of Periostin mRNA is a result of TGS directed by the sasRNA as6 guide, as has been observed previously with other sasRNA targeted loci (Weinberg et al. 2006; Ackley et al. 2013; Morris 2009c), an in-depth investigation into the nuclear expression of *Periostin* was conducted. Nuclear run-on analysis of as6 treated cultures indicated that the observed suppression was transcriptional in nature (**Figure 2-2a**), suggesting that as6 targeting of the *Periostin* 5' UTR is sufficient for modulating the transcriptional activity of *Periostin* in PC3 cells. SasRNA-directed TGS has been observed to result in transcriptional and epigenetic changes at the sasRNA target locus, that ultimately result in a loss of active forms of RNA polymerase II (RNAPII) and a gain of DNA methyltransferase 3a (DNMT3a) at the targeted promoter (reviewed in (Turner and Morris 2010; Morris 2009c)). To determine if as6 treatment affects the localisation of RNAPII or DNMT3a to the *Periostin* promoter a chromatin-immunoprecipitation (ChIP) was performed in as6 treated PC3 cells. ChIP pull-down for active forms of RNAPII revealed a significant reduction in RNAPII and approximately four-fold increased enrichment for DNMT3a at the *Periostin* promoter (**Figure 2-2b**).

Previous studies have observed that TGS can be inhibited by trichostatin A (TSA) and 5'-azacytadine (5' Aza-C), which inhibit the TGS-associated proteins histone deacetylase (HDAC) and DNMT3a respectively (Morris et al. 2004). Treatment of as6 transfected PC3 cells with TSA and 5' Aza-C partially restored *Periostin* gene expression (**Figure 2-2c**). SasRNA directed TGS has been observed to also require transcription at the targeted promoter and a promoter-associated RNA (Han, Kim, and Morris 2007; Napoli et al. 2009b, 2009a). When cells were treated with the drug α amanitin (Kato et al. 2005;

Weinberg et al. 2006), an inhibitor of RNAPII, a complete reversion of mRNA expression is observed (**Figure 2-2d**). These data suggest that TGS of *Periostin* requires RNAPII mediated transcription of the sasRNA guide. Collectively, the observations reported here indicate that as6 is functioning to direct TGS to the *Periostin* promoter in a manner similar to what has been observed previously with other sasRNA targeted genes (Hawkins et al. 2009a; Turner et al. 2012; Turner, De La Cruz, and Morris 2009; Weinberg et al. 2006) (reviewed in (Turner and Morris 2010; Morris 2009c)).

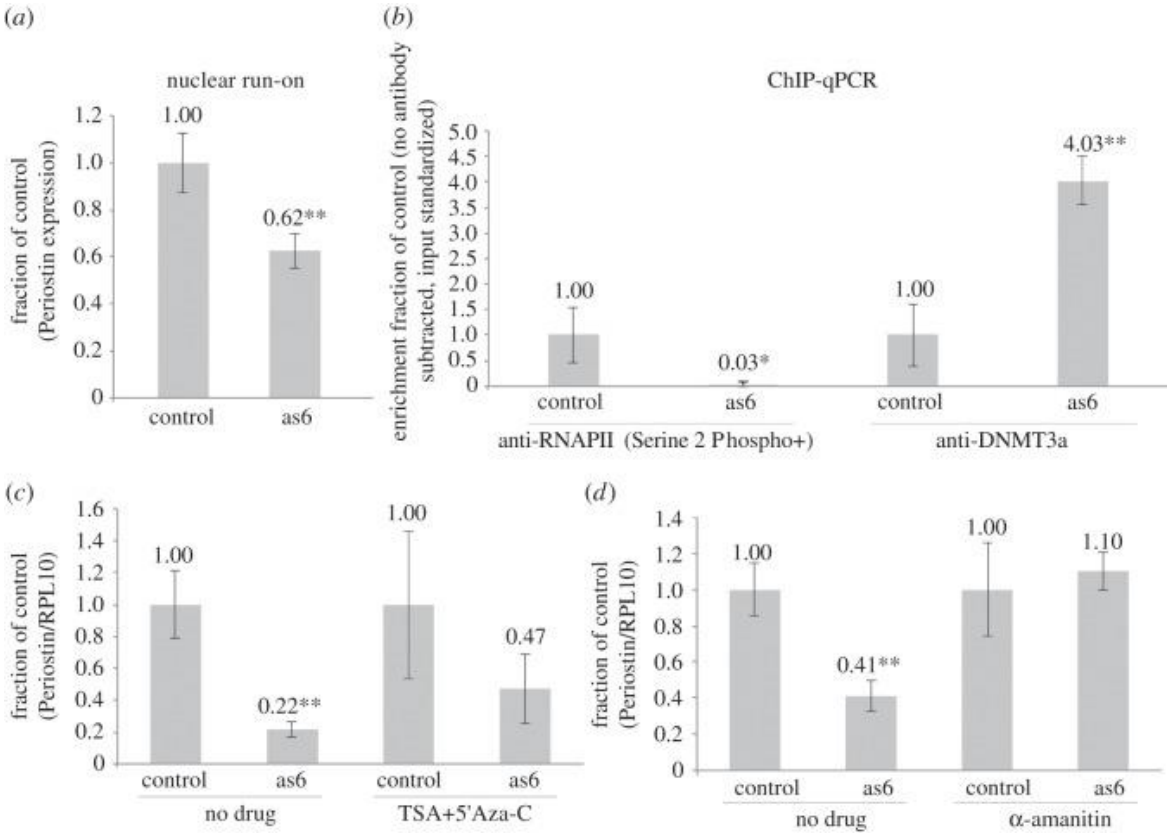


Figure 2-2 *Periostin* TGS in the Nucleus

(**a**) Nuclear run-on analysis of *Periostin* transcripts 72h after transfection with as6, normalised to RPL10 (n=3). (**b**) ChIP of RNAPII and DNMT3a at the *Periostin* promoter, normalised to standard curve of genomic PC3 DNA. Samples are *Periostin* transcripts 72h after transfection with as6 (n=3). (**c**) Drug treatment using both TSA and Aza C. Drugs added every 24 hours to cultures treated as described above (n=3), normalised to RPL10. (**d**) Drug treatment using α amanitin. Drug added once as described above (n=3). All experiments performed in PC3 cells. Throughout figure, * $p<0.05$, ** $p < 0.01$ by two-tailed t-test. Calculations relative to parent U6M2 plasmid. Error bars indicate s.e.m.

2.2.3.2 as6 mediated phenotypes

In an effort to determine if the sasRNA as6 functionally modulates the metastatic ability of tumour cells, we set out to investigate the observable phenotypic effects of as6 when introduced to tumour cells *in vitro*. During the generation of stable HEK293 and PC3 cell lines expressing the as6 sasRNA using Geneticin (G418) selection, it was discovered over a 6-week period that those cells which survived the selection process in both cell lines were unable to proliferate. As such, the generation of stable cell lines was not feasible, possibly due to the suppression of *Periostin* by as6. To explore this notion further cell counts in as6 treated vs untreated PC3 cells were followed over a 72-hour period. Cultures treated with as6 demonstrated a significant reduction in exponential growth (**Figure 2-3a**). Rather than a continued rate of exponential expansion, cell growth was observed to level off and undergo a more linear growth rate after only 3 days of as6 targeting. Notably, by the third day, there were 30% more cells in control samples than in those treated with as6, suggesting as6 directed suppression of *Periostin* impairs cell growth.

Periostin is known to be involved in cell metastasis (Wang and Ouyang 2012; Ben et al. 2011). To explore the ability of as6 to functionally modulate cell motility and metastasis a scratch assay was performed on as6 vs. control treated PC3 cells (**Figures 2-3b - d**). Scratch assay was performed as it was less complex and the resources were less expensive and more readily available than for an invasion assay. Examination of PC3 cell growth into a scratch over a 72-hour period demonstrated severely inhibited growth of cells into the scratch in as6 treated cells relative to control cells (**Figures 2-3b, d**). After 3 days, control cells were seen to move and proliferate into the scratch, covering over the gap left by the plastic insert (**Figure 2-3b**). In contrast, the as6 treated cells showed very little movement into the provided space (**Figure 2-3d**). The addition of mitomycin C to as6 treated cells caused the scratch to remain completely intact, with no edge distortion occurring, suggesting that the

slight regrowth seen without mitomycin C in as6 treated cells was due to cell migration rather than proliferation (**Figure 2-3c**). These data suggest that as6 sasRNA directed TGS of *Periostin* appears to be a potent suppressor of cell migration and is equivalent to the suppression of cell division by mitomycin C.

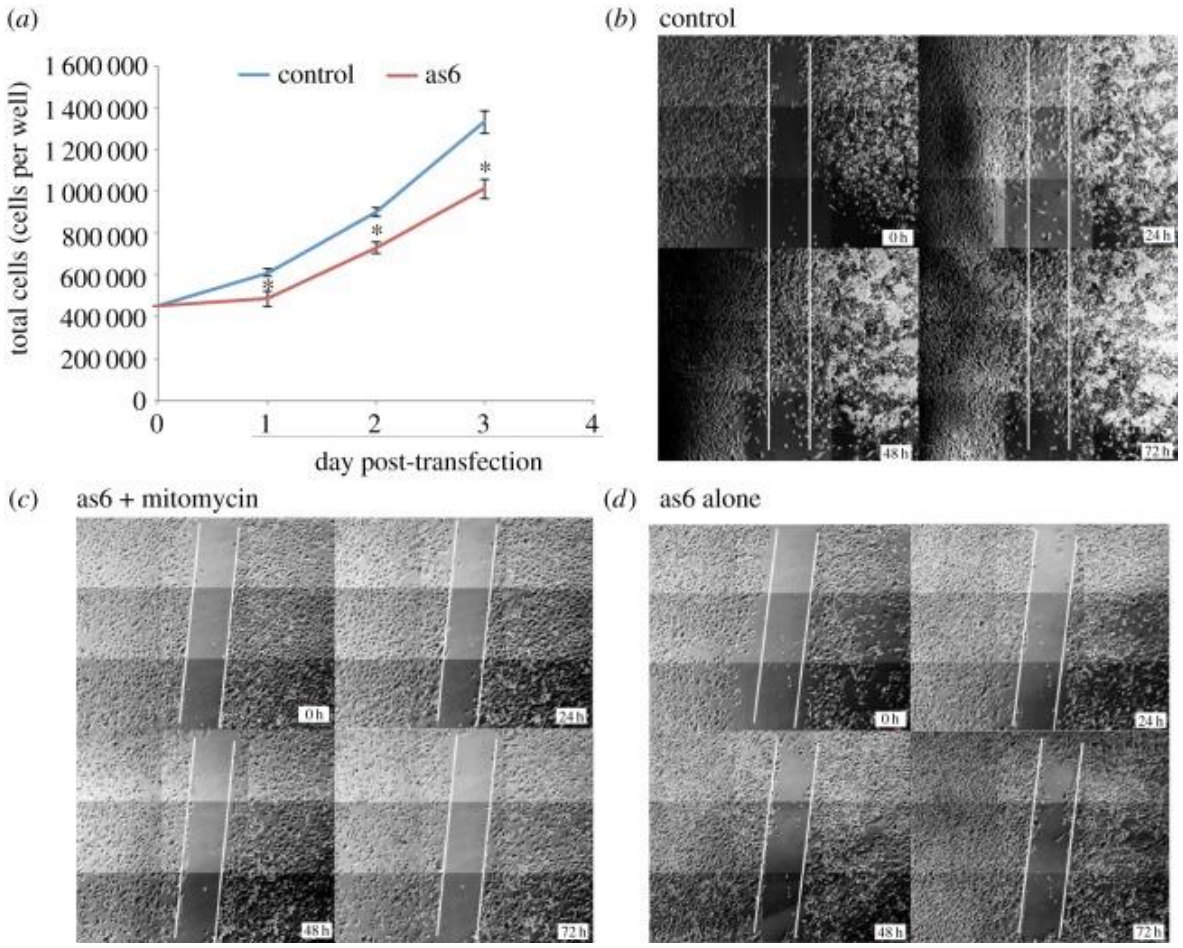


Figure 2-3 Phenotypic Effect of *Periostin* TGS

(**a**) PC3 Cell numbers following as6 treatment. Cells were counted every 24h post-transfection with as6 or parent U6M2 plasmid. * $p<0.05$ by two-tailed t-test. Error bars indicate s.e.m. (**b-d**) Scratch assay 72h after transfection with (**b**) parent U6M2 plasmid alone, (**c**) as6 expressing U6M2 treated with Mitomycin C and (**d**) as6 expressing U6M2 plasmid alone. Cells were photographed at times 0, 24, 48 and 72 hours post-transfection. The white lines indicate initial scratch.

2.2.4 Discussion

Recent studies have presented the concept of endogenous regulatory mechanisms directed by long non-coding RNA's (lncRNA's) in human cells which are responsible for altering the epigenetic state of a target locus, thus affecting the transcriptional expression of a protein coding gene (Morris 2009b; Morris and Mattick 2014a; Turner and Morris 2010). This mechanism can be usurped when a sasRNA guide is introduced into the cell, bypassing endogenous lncRNA regulation of the target locus and inducing TGS (**Figure 2-1a**) (Weinberg et al. 2006; Ackley et al. 2013; Morris 2009c). RNA directed TGS has been observed in human cells before and functions by the sasRNA binding to target low-copy promoter associated transcripts upstream or overlapping the sasRNA targeted promoter (Han, Kim, and Morris 2007; Napoli et al. 2009b, 2009a) resulting ultimately in the recruitment of epigenetic remodelling complexes to the selected site of interest (Morris 2009c; Han, Kim, and Morris 2007) and gene silencing. The advantage of sasRNA directed TGS over other RNA based silencing mechanisms such as RNAi, is that TGS can be long-lasting and heritable. Only a short duration of sasRNA targeting to a gene promoter is required and once DNA methylation is recruited the targeted gene can stay repressed indefinitely (Hawkins et al. 2009a). The data presented here suggests that as6 sasRNA is capable of inducing TGS at the 5'UTR/promoter of *Periostin* and that this targeting has distinct phenotypic effects on cell motility. The increase of DNMT3a occupancy at the *Periostin* promoter is a direct result of as6 activating TGS, leading to methylation and ultimately silencing of *Periostin* (**Figure 2-2a**). Collectively, the observations presented here suggest that as6 has the potential to function as an anti-metastatic molecule that could be highly useful in cancer treatments, although the effect of as6 on cell metastasis remains to be determined and will require further *in vivo* studies.

Previous observations demonstrated a correlation between high expression of *Periostin* and poor cancer patient prognosis (Ben et al. 2009; Choi et al. 2011; Forsti et al. 2007; Nuzzo et al. 2012; Tsunoda et al. 2009). This is presumably due to *Periostin*'s ability to regulate tumour cell invasion and metastasis, as tumour metastasis is associated with the highest rate of mortality for cancer patients (Wang and Ouyang 2012; Ben et al. 2011). Suppression of *Periostin* gene expression *in vitro* via indirect methods (Deraz et al. 2011; Kudo et al. 2006; Malanchi and Huelsken 2009), and others directly through RNAi (Sun, Zhao, et al. 2011; Liu and Liu 2011; Xu et al. 2012), demonstrated a direct correlation between *Periostin* expression and the ability of cells to undergo migration and cell division. While previous studies using RNAi to suppress *Periostin* clearly demonstrated a loss of cell metastasis, the observed effects are only transient as the targeting of *Periostin* was at the post-transcriptional level (Sun, Zhao, et al. 2011; Liu and Liu 2011; Xu et al. 2012). The data presented here suggests that *Periostin* is susceptible to sasRNA directed TGS and that it may be possible to stably suppress *Periostin* (Morris 2009c, 2009b).

Secondary tumours, those that metastasis to new locations, are incapable of surviving without the presence of Periostin (Malanchi et al. 2012). This matches with phenotypic data from the scratch assay and cell count, as well as the inability to generate stable cell lines, as as6 expressing cells would not proliferate and showed very limited ability to migrate, yet they did not undergo apoptosis. Periostin is also required to initiate alteration of the extracellular matrix, creating a niche environment that allows tumour invasion and proliferation (Kim et al. 2005; Malanchi et al. 2012). The observed inability of the as6 treated cells to divide and migrate suggests the involvement of Periostin in rearranging the extracellular matrix and providing an environment conducive to metastasis. When suppression of *Periostin* is induced by TGS, there is no alteration of the surrounding environment to favour cell invasion.

While the TGS induced by as6 was shown to be an effective inhibitor of metastatic and invasive potential of tumour cells, an effective method for introduction of the sasRNA into human patients and specifically to the tumours is required before it can be used to treat cancer patients. New delivery mechanisms that are targeted specifically to diseased cells are required if future treatment employing sasRNA directed TGS becomes a viable option. However, once the sasRNAs reach their targets they have the distinct potential to provide stable epigenetic modifications to the gene promoter of interest and long-term heritable epigenetic silencing.

2.3 Supplemental material that accompanied the manuscript

2.3.1 Materials and methods

2.3.1.1 Cell Culture

PC3 cells were cultured at 37°C in 5% CO_2 using DMEM supplemented with 1% penicillin/streptomycin and 10% FBS.

2.3.1.2 sasRNA's Expressing Constructs

Small antisense RNA (sasRNA) sequences were generated and subsequently cloned in our laboratory into the U6M2 plasmid via *BGl*II and *Kpn*I restriction sites and validated using qRT-PCR as previously described (Ackley et al. 2013). Plasmid digestion was validated by PCR. The sasRNA expressing constructs were transfected into PC3 cells using Lipofectamine 2000, using a ratio of 1µg/10^6 cells (Life Technologies, Carlsbad, CA, USA).

2.3.1.3 RNA Extraction & cDNA Synthesis

RNA extraction was performed using an RNeasy Mini Kit (Qiagen, Redwood City, CA, USA) and subsequently DNase treated (TURBO DNase, Life Technologies Carlsbad, CA, USA). Complimentary DNA (cDNA) was synthesised using 1µg cellular RNA in conjunction with M-MLV Reverse Transcriptase (Sigma/Aldrich, St Louis, MO, USA), dNTP's (New England Biolabs, Ipswitch, USA) and random hexamer (Sigma/Aldrich, St Louis, MO, USA).

2.3.1.4 qRT-PCR

Quantitative RT-PCR (qRT-PCR) analysis was performed using KAPA SYBR FAST qPCR Master Mix (KAPA Biosystems, Woburn, MA, USA) and the desired primer set and cDNA. Plate was placed in ViiA 7 Real-Time PCR system (Life Technologies, Carlsbad,

CA, USA). Cycling conditions were set up as follows: 95°C for 2 mins, then 40 cycles of 95°C for 3s and 60°C for 30s.

2.3.1.5 Stable Cell Lines

Stable cell lines were generated using the antibiotic Geneticin (800µg/ml)(Life Technologies, Carlsbad, CA, USA). PC3 and HEK293 cells were transfected with the desired plasmid, which had undergone linearisation via *Sca*I enzyme digestion. Linearisation was performed to improve the probability of stable plasmid integration into the genome, as linear DNA integrates more efficiently than circular DNA. The selection process was maintained for six weeks, and successful uptake of the plasmid resulted in the continued survival of the cell.

2.3.1.6 Nuclear Run-On

PC3 cells were transfected with either as6 expressing or parent U6M2 plasmid. Cell pellets were collected 72 hours later, followed by cell lysis (10 mM Tris- HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.5% NP-4) and subsequent resuspension (50 mM Tris-HCl, pH 8.3, 0.1 mM EDTA, 40% glycerol, 5 mM MgCl2). Samples were then flash frozen in liquid nitrogen. Upon thawing, in vitro transcription was performed using 50 µl glycerol suspension and 60 µl 2x transcription buffer (20 mM Tris-HCl, pH 8, 5 mM MgCl2, 300 mM KCl, 4 mM DTT, 2 mM ATP, 2 mM CTP, 2 mM GTP, 1 mM biotin-16-AA-5′-UTP, 100 U RNaseOUT (Invitrogen)). Samples were incubated for 45 mins at 30°C before being DNase treated. Nuclei were lysed (50 mM Tris-HCl, pH 7.4, 5% SDS, 0.125 M EDTA) and Proteinase K treated. RNA extraction was performed, and biotinylated RNA pulled down using Dynabeads M-280 Streptavidin (Life Technologies, Carlsbad, CA, USA). Beads were resuspsended in H_2O, allowing cDNA synthesis directly off the beads and subsequent qRT-PCR analysis.

2.3.1.7 Chromatin Immunoprecipitation

As in nuclear run-on, cell pellets were collected 72 hours following transfection, resuspended in PBS, cross-linked with 1% formaldehyde and quenched with 0.125M glycine. Cell lysis (5 mM PIPES, pH 8, 85 mM KCl, 0.5 % NP40) was followed by nuclei lysis (50 mM TRIS-HCl, pH 8, 10 mM EDTA, 1% SDS) and sonication with a Q700 Sonicator (Qsonica). Samples were incubated at 4°C overnight with the desired antibody. Dynabeads Protein G (Life Technologies, Carlsbad, CA, USA) were used to pull down desired antibody, following wash steps in high salt twice (20 mM Tris-HCl, pH 8.1, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), lithium chloride (10 mM Tris-HCl, pH 8.1, 0.25 M LiCl, 1 mM EDTA, 1% sodium deoxycholate, 1% NP40) and TE buffer twice (10 mM Tris-HCl, 1 mM EDTA). DNA/protein is eluted off the beads with 200µl ChIP elution buffer (100mM $NaHCO_3$, pH 8, 1% SDS) and heated to 65°C for 10 minutes. Beads were removed from samples, which were then reverse cross-linked, RNase A and Proteinase K treated. Samples were cleaned up using PCR purification kit (Qiagen, Redwood City, CA, USA) and analysed by qRT-PCR. Antibodies used for ChIP (1 µg/sample) were for RNA Polymerase II (Abcam, Cambridge, MA, USA)(Abcam, cat. no. ab5095) and DNMT3a (Abcam, cat. no. 13888). Determination of relative enrichment is determined by contrasting amplification of a standard curve of genomic PC3 DNA. The bead alone no antibody control samples are subtracted from both the input and CHIPs and the relative enrichment determined as a fraction of input for the respective samples.

2.3.1.8 DNMT3a and HDAC Inhibition

Cells were transfected with desired plasmid, and 12 hours post-transfection both Trichostatin A (TSA) a verified HDAC inhibitor and 5-Azacytadine (5' Aza-C) a verified DNA methyltransferase inhibitor were added to final concentrations 0.05μM and 4μM respectively (Morris et al. 2004). These compounds were then readministered every 24 hours. Seventy-two hours after the initial addition of TSA and 5' Aza-C, RNA extraction was performed, followed by DNase treatment, reverse transcription and qRT-PCR.

2.3.1.9 RNA Pol II Inhibition

Twelve hours following transfection with desired plasmid, α amanitin drug was added to cells to final concentration 0.1 μg/ml (Weinberg et al. 2006). Seventy-two hours after drug addition, RNA extraction was performed, followed by DNase treatment, reverse transcription and qRT-PCR.

2.3.1.10 Polyacrylamide Gel

A 6% polyacrylamide gel was made up with 8 ml 1x Tris/Borate/EDTA (TBE) Buffer, 1.5 ml 40% acrylamide, 100 μl ammonium persulfate (APS) and 10 μl Tetramethylethylenediamine (TEMED). The gel was allowed to set for 20 minutes. qRT-PCR samples from control and as6 treated cells were loaded with gel loading dye onto gel, which was post-stained with ethidium bromide after completion of run and imaged using Geldoc Imaging System (BioRad).

2.3.1.11 Cell Count

Cell pellets were collected 72 hours following transfection and resuspended in 1ml PBS. A 10 μl aliquot was added to 10 μl 0.4% trypan blue. Viable cells were counted using a haemocytometer.

2.3.1.12 Scratch Assay

To a treated 24-well glass bottom plate, sterile micro-inserts were stuck to the bottom of each well. Plate was seeded with PC3 cells starved of FBS. After 24 hours, media was aspirated and 1ml of new DMEM media plus 10% FBS plus 1% penicillin-streptomycin was added. Cells were then transfected with desired plasmid. 12 hours following transfection, Mitomycin C was added to half of the wells from both control and treatment conditions to a final concentration of 20μg/ml. Sterile micro-inserts were then aseptically removed from wells. Photographs were taken of the cells every 15 minutes for 72 hours by a Nikon Eclipse TiE inverted microscope utilising phase contrast microscopy at the Biomedical Imaging Facility (BMIF), UNSW. This microscope contains a stage with a 37°C 5% CO_2 chamber, allowing for maintenance of the cells during cell imaging.

2.3.2 Supplemental data

Table S2-1 Sequences for sasRNA's targeted to *Periostin* promoter

sasRNA	Sequence (5' -- > 3')
as1	GATCTGATGCAGTGTTCCCTCCCACAATTTTTGGTAC
as3	GATCTGAATTTGAAGTTGCCGATGCTTCCTGGCTTTTTGGTAC
as4	GATCTGTTCAGACTCTCAGGTTGATGCTTTTTGGTAC
as5	GATCTGCAACTCTGACATGTATATAAATTCTTTTTGGTAC
as6	GATCTGTTCCCTCCCACAACTCTGACATGTTTTTGGTAC

AGCAGCAATAGTAGAGAAAACATGGGTAAAAAGGGAATCATCTTGAGTCTCTC

CGTTGCAGTTAGTCCCCGAAGAGAACTGGCAGTGGGCTTTGGAGAGCTCAGAA

TTTATATACATGTCAGAGTTGTGGGAGGGAACACTGCATCAACCTGAGAGTCTG

AACTCTTTCCAGGAAGCATCGGCAACTTCAAATTGCCAGCTAAGATTTGTTTTG

GTTAGTTATATTTAAGGAACACATTGAGCTACTTTTCCTTTTCATTATAAAGAA

GGTACTTATTTTTAATCATATGAACAATTGACTCACTGCATGTTTAAGTCTCAGT

TTTATTCCTTTGTATCTTAGAGGGTTTTTAAG

Figure S2-4 Promoter Sequence for *Periostin* Used to Generate sasRNA's

Genomic coordinates chr13:37,598,686-37,599,040

Table S2-2 Oligonucleotides used in study

Oligonucleotide Name	Sequence
POSTN F	GAGCTTTACAACGGGCAAATAC
POSTN R	CTCCCTTGCTTACTCCCTTTC
RPL10 F	CCTCTTTCCCTTCGGTGTG
RPL10 R	AATCTTGGCATCAGGGACAC
POSTNpro F	AGTGGGCTTTGGAGAGCTCAGAAT
POSTNpro R	TTTGAAGTTGCCGATGCTTCCTGG

3 RNA control of transcription in a pseudogene network

Chapter contents:

Page No.

List of tables:

Page No.

List of figures:

Page No.

3.1 The molecular dynamics of long non-coding RNA control of transcription in *PTEN* and its pseudogene

Authors:

Nicholas Lister, Galina Shevchenko, James L. Walshe, Jessica Groen, Per Johnsson, Linda Vidarsdóttir, Dan Grander, Sandro F. Ataide and Kevin V. Morris

Current status of paper:

Published

Date paper accepted for publication or anticipated date of publication:

2 August 2017

Name of Journal:

Proceedings of the National Academy of Sciences U.S.A. 2017; 114(37):9942-9947

Extent to which research is your own:

I performed all ODN based *PTEN* pRNA KD experiments, RIPs for biotin *PTENpg1* asRNA α as well as ChIPs for DNMT3a and methyl-CpG. I also examined the F4R1 variants effect on *PTEN* mRNA expression, as well as the dose curve response of *PTEN* mRNA to increasing amounts of F5R2 variant. I performed all statistical tests and data analysis for these experiments.

Your contribution to writing the paper:

I prepared the first draft of the manuscript, and was subsequently part of large collaborative team effort of editing the manuscript. I prepared figures 1a, c-d, f-g, figure 2 c-I, figures 3 a-d and figures 4 c-d.

Comments:

The following manuscript has been appropriated from a published paper. This manuscript uses the *PTEN* pseudogene regulatory network as a model for examining chromatin interactions with regards to a lncRNA. In the context of this system, the antisense lncRNA

first interacts with another transcript which spans the *PTEN* promoter and contains the 5'-UTR, before recruiting DNMT3a and inducing transcriptional gene silencing. This double-stranded RNA interaction is dependent upon both the structure and sequence of the lncRNA, and that RNA structure may be under a greater evolutionary pressure than sequence conservation.

3.2 Manuscript sent for publication

The molecular dynamics of long non-coding RNA control of transcription; *PTEN* and its pseudogene.

Nicholas Lister[1,^], Galina Shevchenko[2,^], James Walshe[3], Jessica Groen[1], Per Johnsson[4], Linda Vidarsdóttir[5], Dan Grander[5], Sandro Ataide[3], and Kevin V. Morris [2,*]

[1]The University of New South Wales, Biotechnology and Biomedical Sciences, Sydney NSW 2052 Australia

[2]Center for Gene Therapy, City of Hope – Beckman Research Institute and Hematological Malignancy and Stem Cell Transplantation Institute at the City of Hope. 1500 E. Duarte Rd., Duarte, CA, 91010, USA

[3]School of Life and Environmental Sciences, The University Sydney, Sydney NSW 2006 Australia

[4]Ludwig Institute for Cancer Research, Stockholm branch, Sweden

[5]Department of Oncology-Pathology, Cancer Center Karolinska, CCK, R8:03, Karolinska hospital and institute, S-171 76 Stockholm, Sweden

^Contributed equally to this body of work

[*]Corresponding author: kmorris@coh.org

Running Title: *PTENpg1* interactions with the *PTEN* promoter.

Key words: *PTEN*; *PTENpg1*; Epigenetic; DNMT3a; Gene Silencing; pseudogene.

3.2.1 Abstract

RNA has been found to interact with chromatin and modulate gene transcription. In human cells, little is known as to how long non-coding RNAs (lncRNAs) interact with target loci in the context of chromatin. We find here, using the *PTEN* pseudogene as a model system, that antisense lncRNAs interact first with a 5' UTR containing promoter-spanning transcript which is then followed by the recruitment of DNA methyltransferase 3a (DNMT3a); ultimately resulting in the transcriptional and epigenetic control of gene expression. Moreover, we find that the lncRNA and promoter-spanning transcript interaction are based on a combination of structural and sequence components of the antisense lncRNA. These observations suggest, based on this one example, that evolutionary pressures may be placed on RNA structure more so than sequence conservation. Collectively, the observations presented here suggest that a much more complex and vibrant RNA regulatory world may be operative in the regulation of gene expression.

3.2.2 Significance Statement

In recent years non-coding RNA transcripts have been found to interact with genes and modulate their ability to be transcribed and made into protein. Here we uncover many of the mechanistic underpinnings involved in how non-coding RNAs control gene transcription. Notably, we find that non-coding RNA control of gene transcription is based on a combination of structural and sequence components of the non-coding RNA and targeted gene. Collectively, the observations presented here suggest that a much more complex and vibrant RNA regulatory world is operative on gene expression and evolution of the genome.

3.2.3 Introduction

The phosphatase and tensin homologue (*PTEN*) is a protein encoded on chromosome 10 by the *PTEN* gene and is a negative regulator of the oncogenic PI3K-protein kinase B (Akt) pathway. The *PTEN* gene is mutated and epigenetically inactivated in a diverse range of cancers (Steck et al. 1997). This gene is of particular interest as emerging studies have shown that a pseudogene, *PTENpg1*, is actively involved in the regulation of *PTEN*. These observations suggest an underappreciation in the complexity involved in gene regulation. To date, thousands of pseudogenes have been identified in humans, including many disease-associated genes such as TP53 (Pei et al. 2012), BRCA1 (Puget et al. 2002), OCT4 (Hawkins and Morris 2010a; Scarola, Comisso, Pascolo, Chiaradia, Marion, et al. 2015; Suo et al. 2005; Hawkins and Morris 2010b; Scarola, Comisso, Pascolo, Chiaradia, Maria Marion, et al. 2015) and *PTEN* (Dahia et al. 1998). *PTEN* has a single pseudogene in the human genome, *PTENpg1* (also called *PTENp1*, *PTEN*Ψ) that is encoded by chromosome 9 (Teng et al. 1997). *PTENpg1* post-transcriptionally regulates *PTEN* expression by acting as a miRNA sponge to *PTEN*-targeting miRNAs (Poliseno et al. 2010). Recent studies have indicated the presence of antisense RNAs (asRNAs) derived from the PTENpg1 promoter locus (Johnsson, Ackley, et al. 2013b, 2013a). Several different isoforms of this antisense, named α and β, have been identified with transcription arising from the bidirectional *PTENpg1* promoter, and one variant, *PTENpg1* asRNA α, is found to modulate *PTEN* transcription via the recruitment of chromatin modifying complexes EZH2 and DNMT3a (Johnsson, Ackley, et al. 2013b, 2013a). These proteins are actively recruited to the promoter by *PTENpg1* asRNA α and cause chromatin condensation and subsequently a reduction in *PTEN* expression (Johnsson, Ackley, et al. 2013b, 2013a).

It is noteworthy that *PTEN* and PTENpg1 are localised on different chromosomes, and the putative "*in trans"* acting mechanism by which the *PTENpg1* asRNA α interacts with

the *PTEN* promoter has not been determined. We find here that *PTENpg1* asRNA α targeting DNMT3a to the *PTEN* promoter requires transcription of *PTEN*, specifically at the 5'UTR region containing homology to the *PTENpg1* asRNA α transcript, and that this RNA can target the *PTEN* promoter in the absence of DNMT3a.

3.2.4 Results

3.2.4.1 Detection and function of 5' UTR *PTEN* promoter transcripts

Previous studies with small non-coding RNA's (ncRNA's) demonstrated that an expressed low copy transcript spanning the 5' UTR of protein coding genes, designated as a promoter-associated RNA (paRNA), was required for small ncRNA directed epigenetic regulation in human cells (Han, Kim, and Morris 2007; Napoli et al. 2009b; Roberts et al. 2012; Napoli et al. 2009a). Transcriptomic data suggests that there are several expressed sequence tags (ESTs) spanning the *PTEN* promoter (including BG772190 (**Figure 3-1A, Table S3-1**) and DA005202; DA676942; and CN413383). To detect the presence of these transcripts at the *PTEN* promoter, and to examine to what extent that they may play a role in *PTENpg1* asRNA α regulation of *PTEN* transcription, strand specific directional RT and PCR was performed on HeLa total RNA using the PTENproR1 reverse primer only (**Figure S3-5A**). This assay allows for any RNA transcripts spanning the *PTEN* promoter in a sense orientation to be discerned. Notably, a product was observed spanning this region of *PTEN* (**Figure 3-1B**). Unfortunately, it is virtually impossible to disentangle to what extent this *PTEN* promoter-spanning transcript (**Figure 3-1B**) is either a unique low-copy transcript or the 5' UTR of *PTEN* mRNAs (**Figure 3-1A**).

Next, to interrogate to what extent the *PTEN* 5' UTR containing transcripts (paRNA/5'UTR) are involved in *PTENpg1* asRNA α regulation of *PTEN*, suppression of this transcript was carried out using single stranded antisense phosphorothioate oligodeoxynucleotides (ODN) (**Figure S3-5B**) (Eckstein 2000). ODNs allow strand specific interactions to be targeted and blocked, thereby disrupting only the sense-stranded paRNA/5'UTR for *PTEN* (Han, Kim, and Morris 2007; Napoli et al. 2009b; Johnsson, Ackley, et al. 2013b; Napoli, Lemieux, and Jorgensen 1990; Napoli et al. 2009a; Johnsson, Ackley, et al. 2013a). We observed here that targeting the paRNA/5'UTR with ODN2

resulted in suppression of the *PTEN* promoter-associated 5' UTR transcripts (**Figure 3-1C**), but had little to no effect on downstream *PTEN* mRNA expression (**Figure 3-1D**), suggesting that there may be unique 5' UTR associated transcripts overlapping the *PTEN* promoter, similar to previous observations with small RNA targeted transcriptional regulatory mechanisms (Han, Kim, and Morris 2007; Napoli et al. 2009b).

To distinguish whether ODN2 was targeting the *PTEN* gene or a unique 5' UTR associated transcript, we utilized biotin labelled ODN2 to immunoprecipitate (IP) ODN2 bound nucleic acids. By using RNAse A and RNAse H treatment, we determined that ODN2 binds a *PTEN* 5' UTR associated transcript, which interestingly extends the full-length of the *PTEN* mRNA (**Figure 3-1E**), similar to previous observations involved in the mechanism of small non-coding RNA transcriptional gene silencing (Napoli et al. 2009b; Han, Kim, and Morris 2007; Napoli et al. 2009a). Notably, although ODN2 also appeared to bind to the *PTENpg1* transcript (**Figure 3-1E**), this interaction did not appear to affect *PTENpg1* sense and antisense expression (**Figure S3-6**), suggesting that ODN2 interacts with a full-length 5' UTR containing PTEN transcript, which we refer to as paRNA/5'UTR.

The pathway of *PTENpg1* regulation of *PTEN* involves the recruitment of DNMT3a by exon 1 of the *PTENpg1* asRNA α transcript to the *PTEN* promoter (Johnsson, Ackley, et al. 2013b, 2013a). We find here that ODN2 targeting of the *PTEN* paRNA/5'UTR results in a loss of DNMT3a at the *PTEN* promoter (**Figure 3-1F**) and a reduction in CpG methylation at one locus in the *PTEN* promoter (**Figure 3-1G**). This effect of ODN2 on *PTEN* appeared to be the result of ODN2 blocking the activity of the *PTENpg1* asRNA α exon one. Blocking this transcript results in preventing the transcriptional regulation of the *PTEN* promoter by *PTENpg1* asRNA α, as ODN2 treatment in stable *PTENpg1* asRNA α exon 1 over-expressing cells demonstrated increased expression of unspliced forms of *PTEN* (**Figure 3-1H**). Collectively, these observations along with others (Johnsson, Ackley, et al. 2013b), suggest

that *PTEN* promoter-associated 5' UTR containing transcripts are required for *PTENpg1* asRNA α and DNMT3a based epigenetic regulation of *PTEN*.

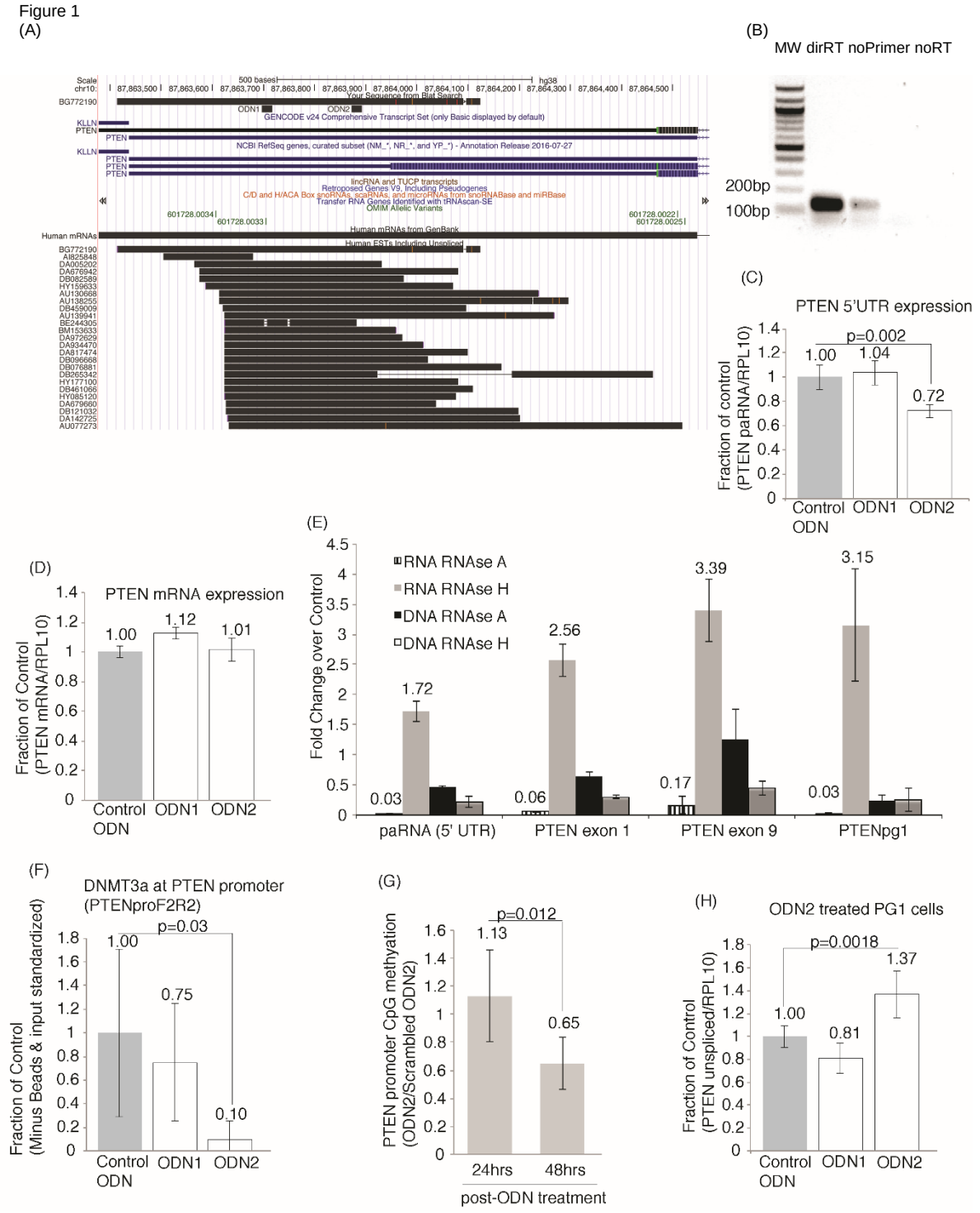


Figure 3-1 The *PTEN* promoter transcriptional landscape

(**A**) A screen shot from the UCSC genome browser depicting several expressed sequence tags (ESTs) at the *PTEN* promoter/5' UTR. EST BG772190 is also shown, which directly overlaps the *PTENpg1* asRNA exon 1 target site (Johnsson, Ackley, et al. 2013b, 2013a). (**B**) Detection of a paRNA (e.g. EST BG772190) by directional RT and PCR were run on a 2% agarose gel. A 100 bp ladder is shown on the left. PCR products are HeLa directional RT (dirRT); HeLa no primer (noPrimer); and no template RT (noRT). (**C-D**) The effects of antisense phosphorothioate oligodeoxynucleotide targeting of the *PTEN* pRNA (EST BG772190). (**C**) Expression of the *PTEN* 5'UTR (e.g. paRNA/EST BG772190) and (**D**) *PTEN* mRNA 48 hours post-transfection with ODNs. (**E**) Biotin ODN2 binds predominantly *PTEN* 5'UTR containing transcripts and not the *PTEN* gene. Biotin ODN2 IP was performed on HEK293 cells and RNAse A vs. RNAse H sensitivity determined. (**F**) Effects of ODN treatment on DNMT3a localisation to the *PTEN* promoter as determined by CHIP 48 hours post-transfection with ODNs. (**G**) *PTEN* promoter CpG methylation as determined by methyl-cytosine specific restriction enzyme McrBc treatment and qPCR in ODN2 and ODN2 Scrambled Control treated HEK293 cells 24 and 48 hours post-transfection. The averages of triplicate treated cultures are shown and the p value from a paired T-test, a single representative experiment is shown. (**H**) The effects of ODN treatment on expression of unspliced variants of *PTEN* in stable HEK293 cells that over-express *PTENpg1* asRNA exon1 (PG1 cells). *PTEN* RNA expression was determined by qRT-PCR with *PTEN* F3/R3 primers; which specifically detect unspliced forms of *PTEN*. For C-D and F and H triplicate treated samples are shown as a fraction of the control ODN ± standard error of mean (SEM) after normalization to RPL10 and for E duplicates with the ranges shown from a single representative experiment. P values from a paired two-tailed t-test are also shown.

3.2.4.2 The *PTEN* paRNA/5'UTR interacts directly with *PTENpg1* asRNA α exon 1

Previous studies with small antisense RNAs (sasRNAs) targeted to gene promoters have demonstrated that the sasRNAs interact directly with a transcript at the promoter (Han, Kim, and Morris 2007; Napoli et al. 2009b, 2009a). The observations presented here (**Figure 3-1**), along with previous studies (Johnsson, Ackley, et al. 2013b), suggest that the long non-coding RNA (lncRNA) *PTENpg1* asRNA α regulates *PTEN* transcription by interactions with a sense-stranded transcript or elongated 5'UTR (paRNA/5'UTR) that essentially spans the *PTEN* promoter. To interrogate whether the *PTEN* promoter-associated transcript interacts directly with *PTENpg1* asRNA α and DNMT3a, an immunoprecipitation of DNMT3a was carried out followed by qRT-PCR with primers specific to each transcript (**Figure 3-2A**). We find using this technique that DNMT3a interacts with both *PTENpg1* asRNA alpha exon 1 and a *PTEN* paRNA/5' UTR promoter-associated transcript (**Figures 3-2A - B**).

To interrogate whether the *PTEN* promoter-associated transcript interacts directly with *PTENpg1* asRNA α, biotin labeled *PTENpg1* asRNA α Exon1 was generated and co-transfected with ODN2 into HEK293 cells, and the effects on biotin labeled *PTENpg1* asRNA α exon1 binding to the *PTEN* promoter determined by RNA immunoprecipitation (RIP) (**Figures 3-2D - I)** (Hawkins and Morris 2010a). ODN2 treatment, relative to controls, abrogated the binding of *PTENpg1* asRNA α exon 1 to the *PTEN* promoter (**Figures 3-2D - F**). To more clearly map the interacting domain of the *PTEN* 5' UTR promoter-associated transcript (paRNA/5'UTR) and *PTENpg1* asRNA α, RNAse A treatment was carried out on the ODN2 and biotin *PTENpg1* asRNA α exon 1 co-transfected HEK293 cells. Of those regions bound to the *PTEN* promoter in the presence of ODN2, one particular region spanning the *PTENpg1* asRNA α exon 1 homologous target region and ~ 150 nucleotides upstream in the *PTEN* promoter (**Figure 3-2G**) appeared to exhibit the most abundant

RNA:RNA interactions, based on RNAse A sensitivity, relative to the other loci assessed (**Figures 3-2H - I**), suggesting that this region (**Figure 3-2C**, PTENproF1/R1) may be where the strongest localisation of those RNA:RNA interactions required for *PTENpg1* asRNA α regulation of *PTEN* transcription occur.

Figure 2

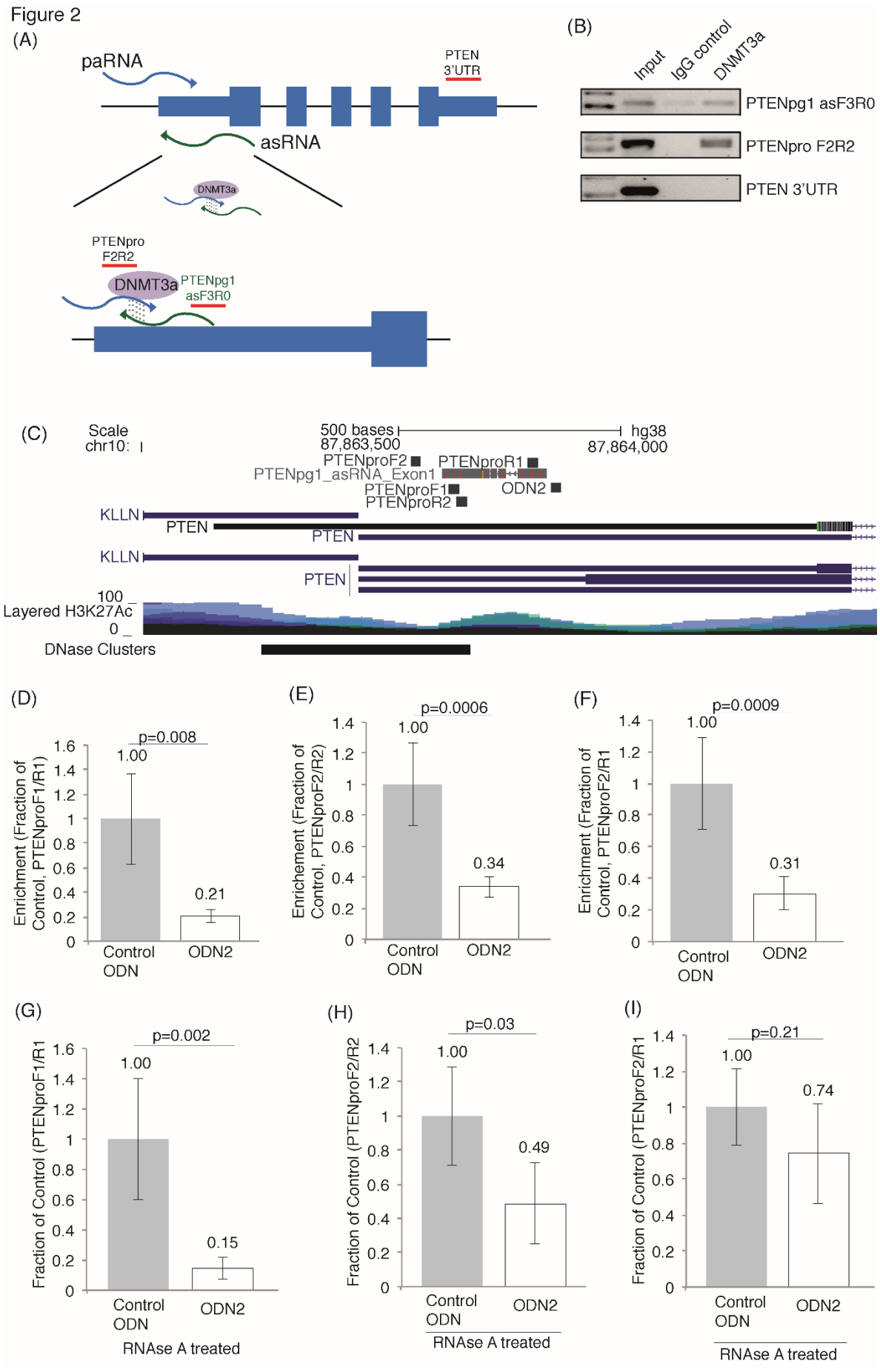

(A)
paRNA
PTEN 3'UTR
asRNA
DNMT3a
PTENpro F2R2
DNMT3a
PTENpg1 asF3R0
(B)
Input
IgG control
DNMT3a
PTENpg1 asF3R0
PTENpro F2R2
PTEN 3'UTR
(C)
Scale
chr10:
500 bases
87,863,500
hg38
87,864,000
PTENproF2
PTENproR1
PTENpg1_asRNA_Exon1
PTENproF1
ODN2
PTENproR2
KLLN
PTEN
PTEN
KLLN
PTEN
100
Layered H3K27Ac
0
DNase Clusters
(D)
p=0.008
1.00
0.21
Enrichment (Fraction of Control, PTENproF1/R1)
Control ODN
ODN2
(E)
p=0.0006
1.00
0.34
Enrichement (Fraction of Control, PTENproF2/R2)
Control ODN
ODN2
(F)
p=0.0009
1.00
0.31
Enrichment (Fraction of Control, PTENproF2/R1
Control ODN
ODN2
(G)
p=0.002
1.00
0.15
Fraction of Control (PTENproF1/R1
Control ODN
ODN2
RNAse A treated
(H)
p=0.03
1.00
0.49
Fraction of Control (PTENproF2/R2
Control ODN
ODN2
RNAse A treated
(I)
p=0.21
1.00
0.74
Fraction of Control (PTENproF2/R1
Control ODN
ODN2
RNAse A treated

Figure 3-2 The effects of ODN2 treatment on PTENpg1 asRNA exon1 binding to and regulation of *PTEN*

(**A**) A schematic is shown depicting the predicted *PTEN* paRNA and *PTENpg1* asRNA exon 1 interacting region relative to *PTEN* along with the RNA:RNA interacting domain and presumed DNMT3a interaction (Johnsson, Ackley, et al. 2013b, 2013a). (**B**) CHIP was carried out for DNMT3a in HEK293 cells and those transcripts associated with DNMT3a were determined by semi-qRT-PCR. DNMT3a was observed to interact with both *PTENpg1* asRNA (asF3R0) and paRNA/EST BG772190 (PTENpro F2R2) loci but not the 3' UTR of *PTEN*. The PTENpro F2/R2 primer set is specific for *PTEN*-5'UTR paRNAs (e.g. EST BG772190), while PTENpg1 asF3R0 is specific for *PTEN* pg1-encoded transcripts. A primer set targeting the *PTEN* 3'UTR is provided as a control. (**C**) A UCSC genome browser screenshot of the *PTENpg1* antisense RNA exon 1 target site in the *PTEN* promoter is shown along with various primers used to interrogate *PTENpg1* asRNA exon 1 interactions with the *PTEN* promoter. (**D**-**I**) RIP analysis of biotin *PTENpg1* asRNA α interactions with the *PTEN* promoter in the presence of ODN2. (**D**-**F**) ODN2 treatment blocks biotin dUTP containing *PTENpg1* asRNA exon1 binding to the *PTEN* promoter. Biotin labeled *PTENpg1* asRNA exon 1 was co-transfected with ODN2 into HEK293 cells and localisation to the *PTEN* promoter determined by CHIP. (**D**) Primer set PTENproF1/R1, (**E**) PTENpro F2/R2 and (**F**) PTENproF2/R1 were used to detect the loss of *PTENpg1* asRNA α exon 1 binding, following ODN2 treatment, to various loci in the *PTEN* promoter. (**G-I**) Mapping of RNA:RNA interacting loci at the *PTEN* promoter. ODN2 treatment results in reduced *PTENpg1* asRNA exon1 interactions with *PTEN* paRNA/EST BG772190 at the (**G**) PTENproF1/R1 and (**H**) PTENproF2/R2 loci but not the (**I**) PTENproF2/R1 locus. For (**D-I**) the averages of triplicate treated cultures are shown with the standard error of mean (SEM) and p values from a paired two-tailed t-test.

3.2.4.3 Determining the mechanistic parameters of *PTENpg1* asRNA α exon 1 interactions with the *PTEN* 5'UTR

Previous observations have indicated *PTENpg1* asRNA α exon 1 binds and directs DNMT3a to the *PTEN* promoter (Johnsson, Ackley, et al. 2013b, 2013a). To determine the parameters and particular region in *PTENpg1* asRNA α exon 1 involved in binding to the *PTEN* promoter, various truncations of *PTENpg1* asRNA α exon 1 were generated as biotin labeled transcripts (**Figure S3-7 and Table S3-2**). These truncated *PTENpg1* asRNA α exon 1 variants (**Figure S3-7A and Table S3-2**) were transfected into HEK293 cells and localisation to the *PTEN* promoter determined by chromatin immunoprecipitation (CHIP). Two truncated fragments appeared to localise to the *PTEN* promoter, F4R1 and F5R2, as well as the control full-length *PTENpg1* asRNA α exon 1 (**Figures 3-3A and S3-7**). However, when these deletion constructs were assessed for their ability to direct DNMT3a to the *PTEN* promoter, only the full-length *PTENpg1* asRNA α exon 1 and F4R1 variant was functionally capable of directing DNMT3a to the *PTEN* promoter (**Figure 3-3B**).

Interestingly, the F4R1 truncated variant was able to repress *PTEN* mRNA expression (**Figure 3-3C**), similar to previous observations with *PTENpg1* asRNA exon 1 (Johnsson, Ackley, et al. 2013b, 2013a), while the F5R2 variant resulted in a dose dependent increase in *PTEN* expression (**Figure 3-3D**). Indeed, both *PTENpg1* asRNA α exon 1 and F4R1 were found to interact directly with DNMT3a/DNMTL-CD *in vitro* (**Figures 3-3E - F and S3-8 - 9**), relative to the control GFP RNA (**Figure 3-3H**) as observed in electrophoretic mobility shift assays (EMSA). Interestingly, the F5R2 variant was also found to bind DNMT3a/DNMTL-CD *in vitro* (**Figure 3-3G**), but unlike the full-length *PTENpg1* asRNA α exon 1 or F4R1 variants, was unable to direct DNMT3a to the *PTEN* promoter (**Figure 3-3B**). This is an interesting observation as the F5R2 variant appears to bind the *PTEN* promoter (**Figure 3-3A**) or DNMT3a (**Figure 3-3G**) but not both (**Figure 3-3B**) suggesting,

based on this single observation, that some antisense lncRNAs may target genes in the absence of DNMT3a or bind DNMT3a and block endogenous recruitment to their intended target. In the case presented here with the *PTEN* locus, the dose dependent over-expression of the F5R2 variant appeared to result in active increases in *PTEN* expression (**Figure 3-3D**).

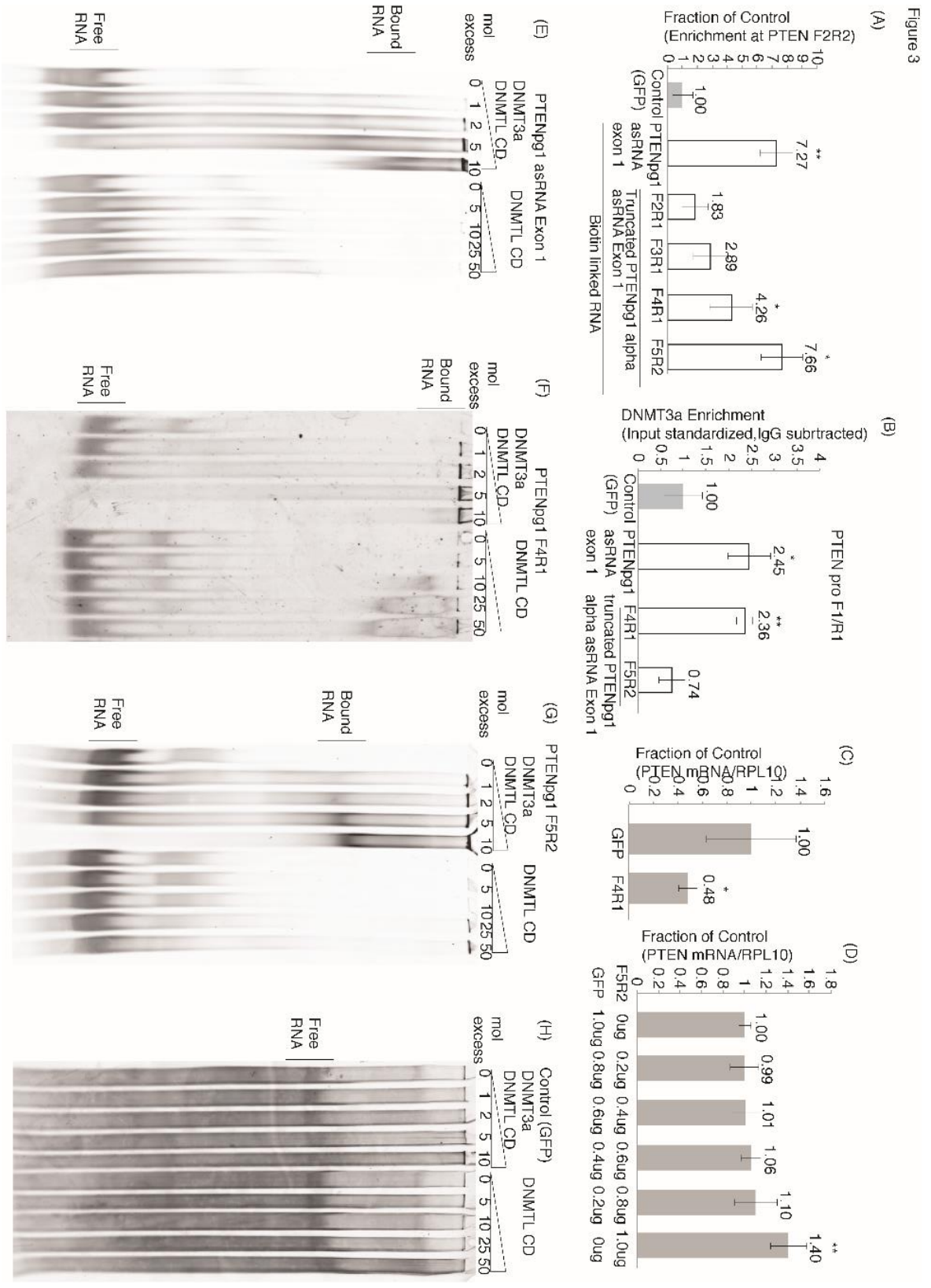


Figure 3
(A)
Fraction of Control
(Enrichment at PTEN F2R2)
Control (GFP)
PTENpg1 asRNA exon 1
Truncated PTENpg1 asRNA Exon 1
PTENpg1 alpha
Biotin linked RNA
F2R1
F3R1
F4R1
F5R2
1.00
7.27
1.83
2.89
4.26
7.66
(B)
DNMT3a Enrichment
(Input standardized, IgG subtracted)
PTEN pro F1/R1
1.00
2.45
2.36
0.74
(C)
Fraction of Control
(PTEN mRNA/RPL10)
GFP
F4R1
1.00
0.48
(D)
Fraction of Control
(PTEN mRNA/RPL10)
GFP 1.0ug 0.8ug 0.6ug 0.4ug 0.2ug 0ug
F5R2 0ug 0.2ug 0.4ug 0.6ug 0.8ug 1.0ug
1.00 0.99 1.01 1.06 1.10 1.40
(E)
PTENpg1 asRNA Exon 1
DNMT3a
DNMTL CD
mol excess
0 1 2 5 10 0 5 10 25 50
Bound RNA
Free RNA
(F)
PTENpg1 F4R1
(G)
PTENpg1 F5R2
(H)
Control (GFP)

Figure 3-3 Characterisation of *PTENpg1* asRNA α exon 1 and truncated variants involvement in the recruitment of DNMT3a to the *PTEN* promoter

(**A**) Enrichment of biotin labelled *PTENpg1* asRNA α full-length and truncated variants by RIP at the *PTEN* promoter (F2/R2 locus) following subtraction of beads alone and standardised to inputs. (**B**) Ability of *PTENpg1* asRNA exon 1 and truncated variants to direct DNMT3a to the *PTEN* promoter as determined by CHIP. (**C**) The *PTEN* F4R1 variant suppresses *PTEN* mRNA expression. Either the control pcDNA3.1-GFP or pcDNA-F4R1 were transfected in triplicate into HEK293 cells and *PTEN* mRNA expression determined 72 hours later by qRT-PCR. (**D**) The *PTEN* F5R2 variant activates *PTEN* expression based on qRT-PCR analysis of transcript expression. Dose dependent transfection of F5R2 activation of *PTEN* mRNA expression was observed in HEK293 cells transfected in triplicate with increasing concentrations of either pcDNA3.1-F5R2 or control pcDNA3.1-GFP. (**E-H**) EMSA analysis of (**E**) *PTENpg1* asRNA exon 1, (**F**) Truncated F4R1, (**G**) Truncated F5R2, and (**H**) Control GFP binding to DNMT3a in the presence of DNMTL-CD. The reconstituted system Dnmt3a/DnmtL-CD *in vitro* requires the presence of DNMTL-CD, which alone is unable to bind RNA (even at 50x molecular excess) indicating that the observed RNA binding in the DNMT3a/DNMTL-CD complex comes from DNMT3a. For A-B the averages of triplicate transfected HEK293 cells from a single representative experiment are shown with the standard error of the mean and p values from a two-sided T-test (*$p<0.05$ or ** for $p<0.01$). For D (**) a one-way ANOVA was carried out with the *f*-ratio value of 10.38725 and the *p*-value is 0.00913.

To interrogate this notion further and determine the requirement of an observed major loop (5'-ACAUUCAAACUUCCAUCACGGC-3') that was found in both the *PTENpg1* asRNA exon 1 (**Figure S3-7B**) and F4R1 variant (**Figures 3-4A and S3-7C**) but not in the F5R2 variant (**Figure S3-7D**) in binding to the *PTEN* promoter, we generated several truncations of F4R1 (**Figure 3-4A, Table S3-3**) and determined their respective ability, along with the full-length F4R1 control, to bind the *PTEN* promoter in the absence of DNMT3a. Interestingly, only the controls full-length F4R1 and *PTENpg1* asRNA exon 1 were able to bind the *PTEN* promoter in the absence of DNMT3a (**Figure 3-4B**), suggesting that the major loop observed in both the *PTENpg1* asRNA exon 1 and the F4R1 variant is required for *PTENpg1* asRNA targeting of the *PTEN* promoter in the absence of DNMT3a (**Figures 3-4C - D**).

Figure 4

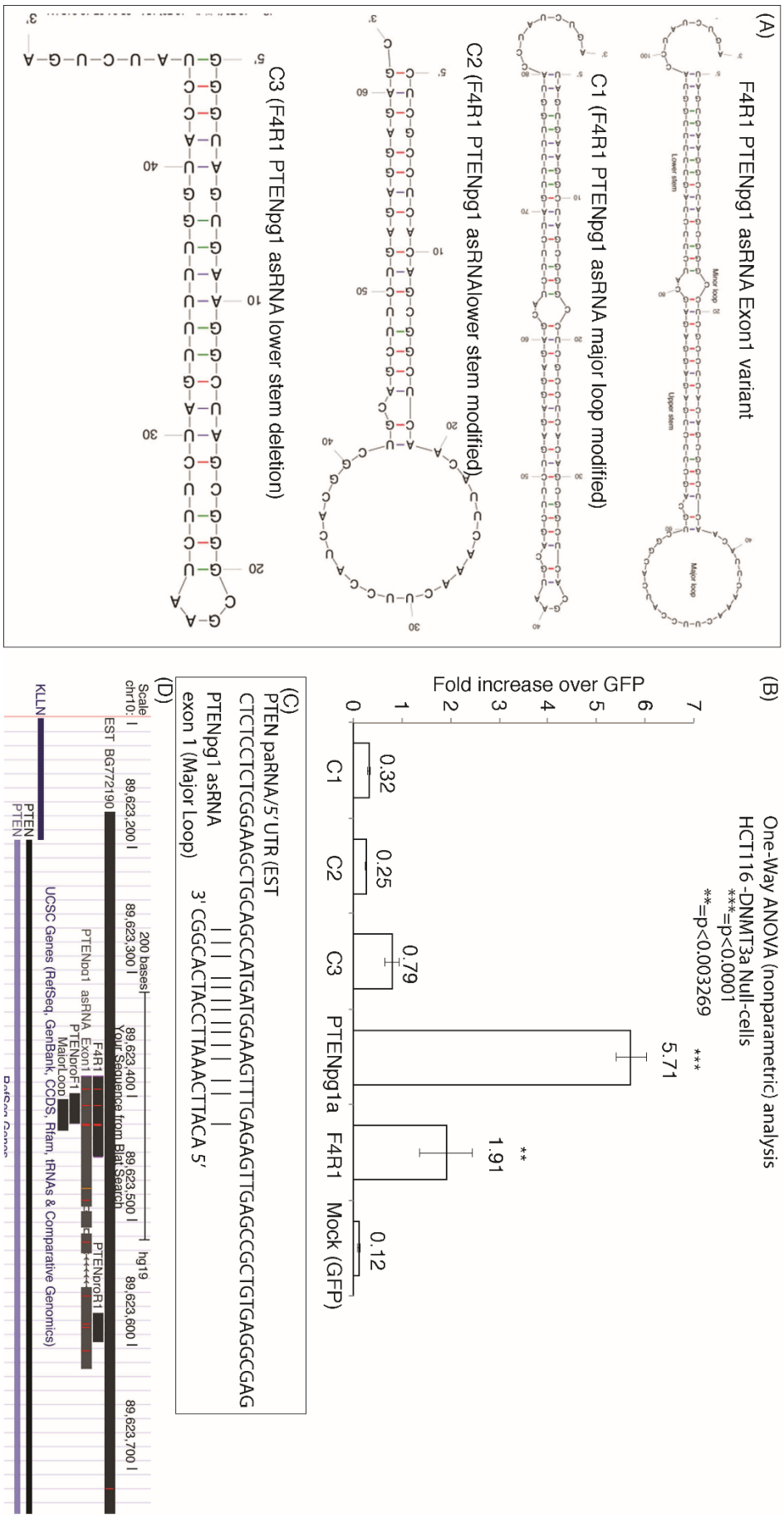

(A)
F4R1 PTENpg1 asRNA Exon1 variant
Lower stem
Minor loop
Upper stem
Major loop
C1 (F4R1 PTENpg1 asRNA major loop modified)
C2 (F4R1 PTENpg1 asRNAlower stem modified)
C3 (F4R1 PTENpg1 asRNA lower stem deletion)
(B)
One-Way ANOVA (nonparametric) analysis
HCT116 -DNMT3a Null-cells
***=p<0.0001
**=p<0.003269
Fold increase over GFP
0.32
0.25
0.79
5.71
***
1.91
**
0.12
C1
C2
C3
PTENpg1a
F4R1
Mock (GFP)
(C)
PTEN paRNA/5'UTR (EST
CTCTCCTCTCGGAAGCTGCAGCCATGATGGAAGTTTGAGAGTTGAGCCGCTGTGAGGCGAG
PTENpg1 asRNA
exon 1 (Major Loop)
3' CGGCACTACCTTAAACTTACA 5'
(D)
Scale
chr10:
89,623,200
89,623,300
89,623,400
89,623,500
89,623,600
89,623,700
200 bases
hg19
Your Sequence from Blat Search
EST BG772190
PTENpg1 asRNA Exon1
F4R1
PTENproF1
MajorLoop
PTENproR1
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
KLLN
PTEN
PTEN

Figure 3-4 Characterisation of the *PTENpg1* asRNA F4R1 variant functional domain

(**A**) Various truncations of F4R1 were developed and cloned into the pcDNA3.1 expression vector, and screened for binding to the *PTEN* promoter. M-Fold analysis of the predicted RNAs is shown. (**B**) Biotin labelled *PTENpg1* asRNA Exon1, F4R1, and F4R1 variants C1-F4R1 *PTENpg1* asRNA α major loop; C2-F4R1 PTENpg1 asRNA α lower stem; and C3-F4R1 *PTENpg1* as RNA α upper stem and major loop binding to the *PTEN* promoter in the DNMT3a deleted HCT116 cultures. The average of triplicate treated cultures are shown with p values from a One-Way ANOVA (nonparametric) analysis (***=$p<0.0001$ **=$p<0.003269$). (**C**) *PTENpg1* asRNA exon 1 and F4R1 major Loop interacting locus in the *PTEN* paRNA (EST BG772190)/5' UTR of *PTEN*). (**D**) A screen shot from the UCSC genome browser showing the PTENproF1/R1 region where *PTENpg1* asRNA exon 1 and F4R1 interact with EST BG772190. The major loop interacting region is also shown.

3.2.5 Discussion

The observations presented here suggest that *PTEN* 5' UTR promoter-associated transcripts are involved in *PTENpg1* asRNA exon 1 directed epigenetic regulation of *PTEN*. Previous studies have detected promoter-associated transcripts (Di Ruscio et al. 2013; Napoli et al. 2009b), which are required for small antisense RNAs to guide DNMT3a and direct transcriptional gene silencing in human cells (Han, Kim, and Morris 2007; Napoli et al. 2009b; Roberts et al. 2012; Napoli et al. 2009a). These promoter-associated transcripts are thought to be low abundant mRNAs which contain elongated 5' UTRs that can be detected through directional RT and PCR (Han, Kim, and Morris 2007; Napoli et al. 2009b, 2009a) and public deep sequencing (**Figure 3-1A and Table S3-1).**

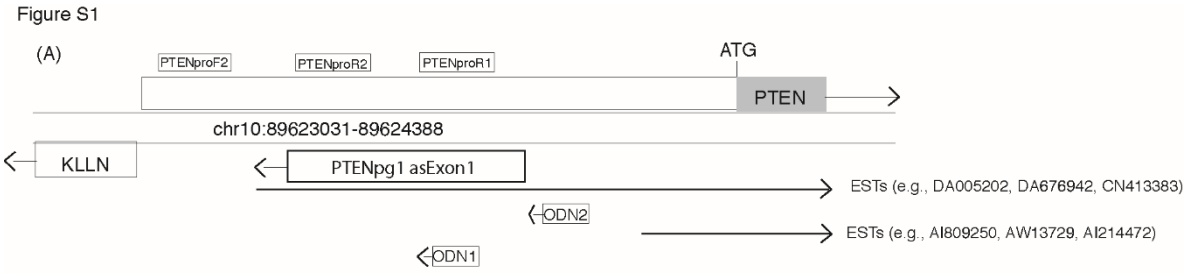
The data presented here juxtaposed with previous observations suggest that a mechanism of action is active in human cells whereby RNA:RNA interactions occur at chromatin to facilitate the recruitment of epigenetic regulatory protein complexes (Morris 2009a). Building on observations that *PTENpg1* antisense RNA α exon 1 is an active transcriptional and epigenetic modulator of *PTEN* (Johnsson, Ackley, et al. 2013b, 2013a), we find here that the localisation of *PTENpg1* asRNA α exon 1 and a truncated variant F4R1 to the *PTEN* promoter requires a *PTEN* promoter-associated RNA (paRNA/5'UTR) and involves a conserved major loop domain (5'-ACAUUCAAACUUCCAUCACGGC-3') to successfully direct DNMT3a to the *PTEN* promoter (**Figure 3-4**). Interesting, through deletion studies one sequence appears to be the main modulator involved in the ability of *PTENpg1* asRNA exon1 or F4R1 to target DNMT3a to the *PTEN* promoter, which was not retained in the defective F5R2 variant. This sequence (**Figure 3-4C**) appears in both the *PTENpg1* asRNA exon 1 and F4R1 variants and maps directly to a region that was observed previously to exhibit high levels of DNMT3a and *PTENpg1* asRNA exon 1 binding (**Figure 3-4D**) (Johnsson, Ackley, et al. 2013b, 2013a).

The observations presented here suggest that an RNA:RNA interaction is involved in *PTENpg1* asRNA exon1 targeting of the *PTEN* promoter and that the conserved domain required for this interaction consists of the major loop domain (5'-ACAUUCAAACUUCCAUCACGGC-3') interacting with the *PTEN* paRNA/5'UTR (**Figures 3-4C - D**). This loop appears to be required for localisation of the *PTENpg1* asRNA exon 1 and F4R1 variant transcripts to the *PTEN* promoter along with the entire stem present in the F4R1. When this loop is altered, as is the case with the F5R2 variant, there appears to be a loss of localisation of the transcript to the *PTEN* promoter, while an ability to interact with DNMT3a remained intact.

Collectively, the dichotomous observations presented here between *PTENpg1* asRNA exon1, F4R1, and the F5R2 variants suggest that the major loop domain (**Figure 3-4A, 3-4C, and S3-7B-C**), in combination with a longer stem found in both *PTENpg1* alpha asRNA exon 1 and the F4R1 variant, interacts directly with DNMT3a to direct transcriptional and epigenetic silencing of *PTEN*. Such observations may support the notion that lncRNAs and their putative evolutionary conservation may be more contingent on a combination of both structure and sequence (Johnsson et al. 2014). Collectively, the observations presented here expand our understanding of endogenous lncRNA networks in human cells and suggest that RNA:RNA interactions, particularly at gene promoters, may be mechanistically relevant in lncRNA regulation of protein-coding gene expression. An understanding of this emerging mode of gene and epigenetic regulation could prove useful in the development of targeted therapeutics to disrupt or augment transcriptional regulatory networks in humans.

3.3 Supplemental material that accompanied the manuscript

3.3.1 Supplemental data



(B)

ODN	Target	Sequence
ODN 1	PTEN paRNA	5'-GTC TCT CAT CTC CCT CGC CT-3'
ODN 2	PTEN paRNA	5'-GCT TCC ACC TTC CCT TTC AG-3'
CTRL ODN	miR367 control	5'-ACT GAC CTT TGG ATG GTG CT-3'

Figure S3-5 Location of directional RT and PCR primers and antisense ODNs relative to the *PTEN* promoter

(**A**) A schematic of the *PTEN* promoter is shown depicting the *PTENpg1* asRNA exon1 binding site and location of antisense oligodeoxinucleotide target sites (ODNs) as well as RT and PCR primers used in directional RT analysis of the *PTEN* associated paRNA. (**B**) The sequence of the ODNs and control ODN (miRN367, (Omoto et al. 2004)) used in the presented study.

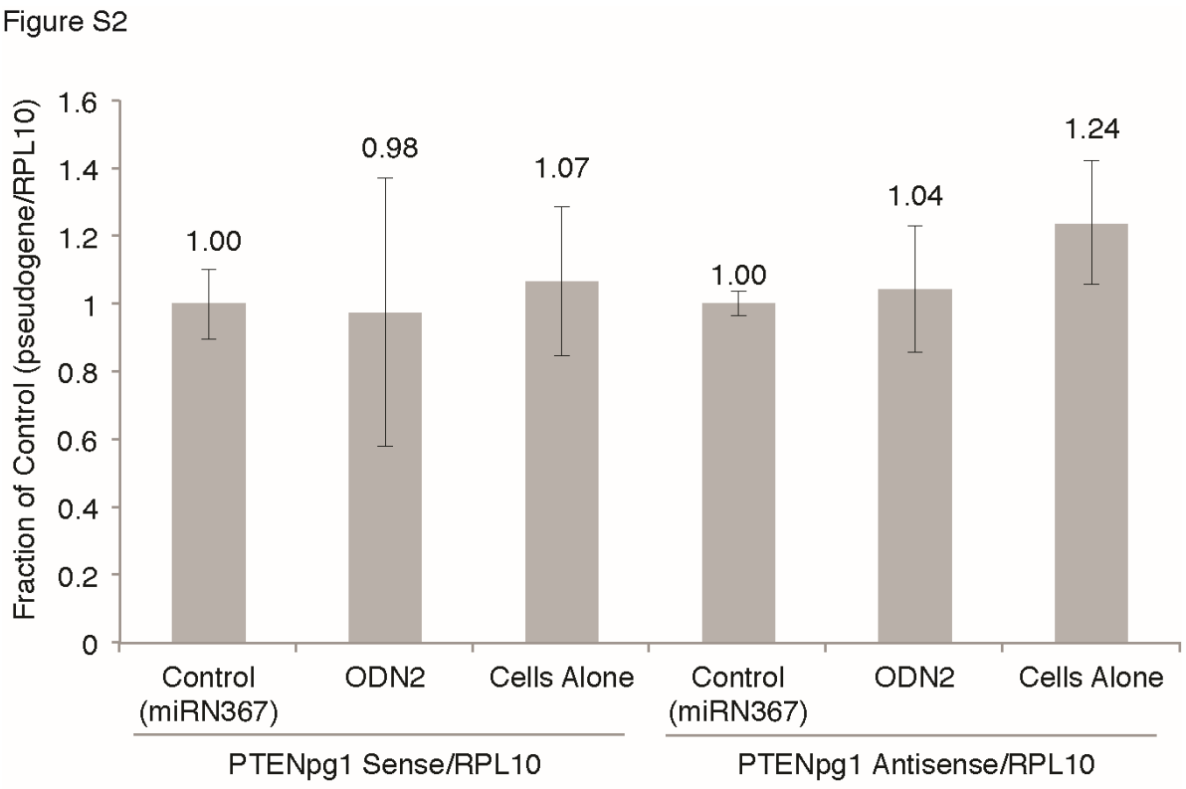


Figure S3-6 *PTENpg1* and *PTENpg1* asRNA expression following ODN2 treatment

HEK293 cells were transfected in triplicate with ODN2, Control ODN miRN37 target site (Omoto and Fujii 2005; Omoto et al. 2004). The averages of triplicate treated cultures are shown with the standard deviations.

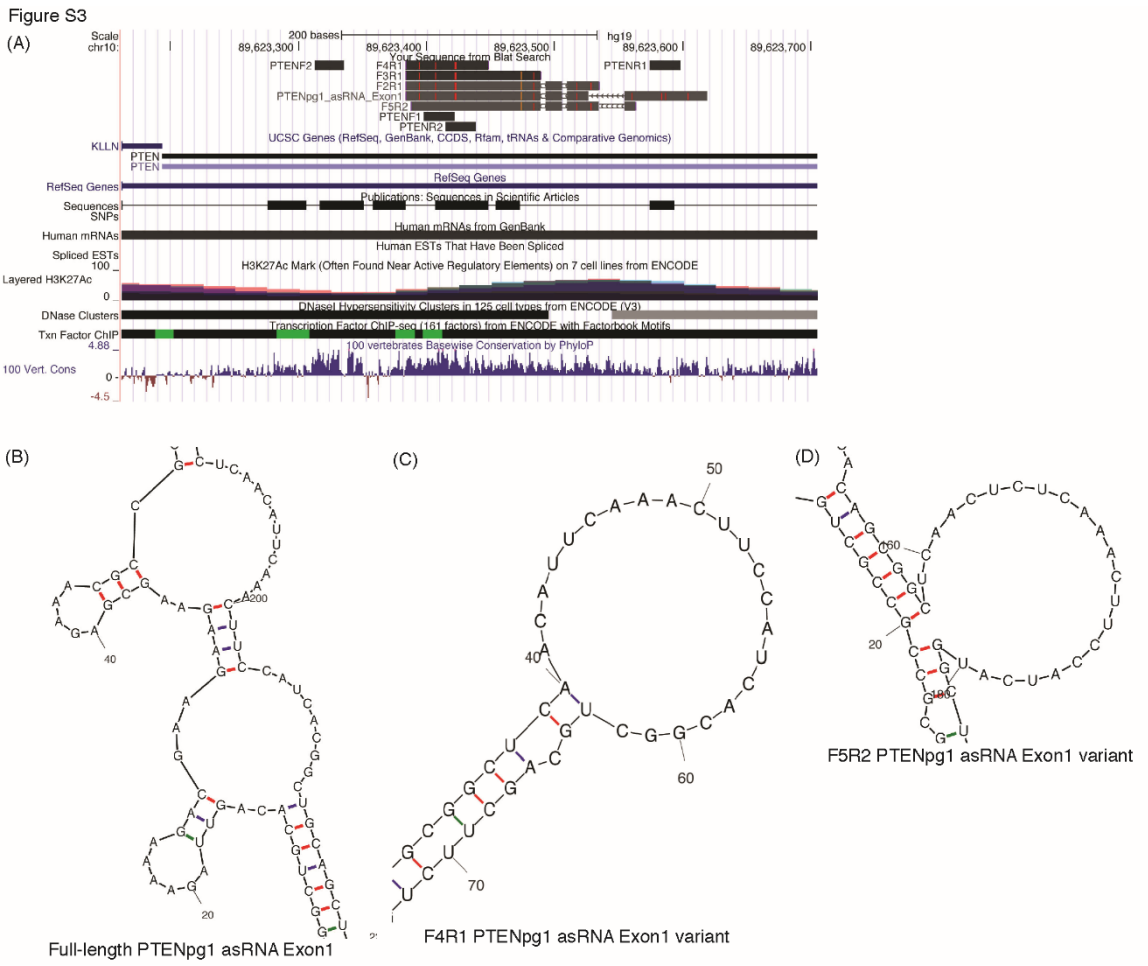


Figure S3-7 Analysis of the *PTENpg1* asRNA alpha exon 1 and various truncations of *PTENpg1* asRNA exon1 target sites in the *PTEN* promoter

(**A**) A screen shot from UCSC genome browser is shown depicting the *PTEN* promoter and homologous target sites for *PTENpg1* asRNA exon1 and the various truncations of *PTENpg1* asRNAs assessed. (**B-D**) The predicted DNMT3a binding bulge in both full-length and F4R1 variant of *PTENpg1* asRNA Exon 1 as determined from M-Fold. The conserved sequence that appears required for DNMT3a interactions with *PTENpg1* asRNA exon 1 (**B**) at the *PTEN* promoter (ACAUUCAAAC UUCCAUCACGGC) is shown in *PTENpg1* asRNA exon 1 and F4R1 (**C**) but lost in F5R2 (**D**).

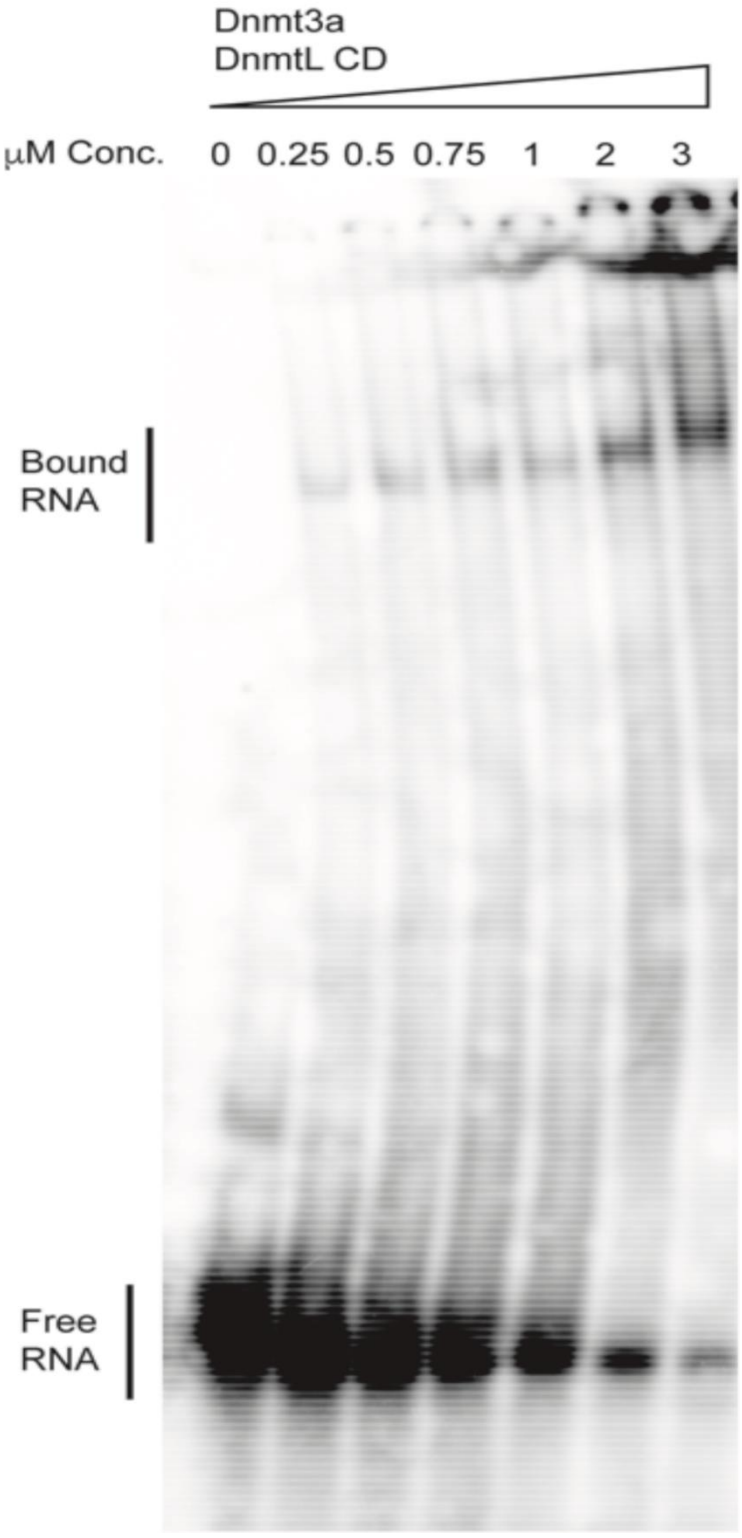


Figure S3-8 EMSA analysis of *PTENpg1* asRNA truncated F4R1 binding to DNMT3a in the presence of DNMTL-CD

Protein concentration is indicated above and the free RNA and bound RNA-DNMT3a are indicated.

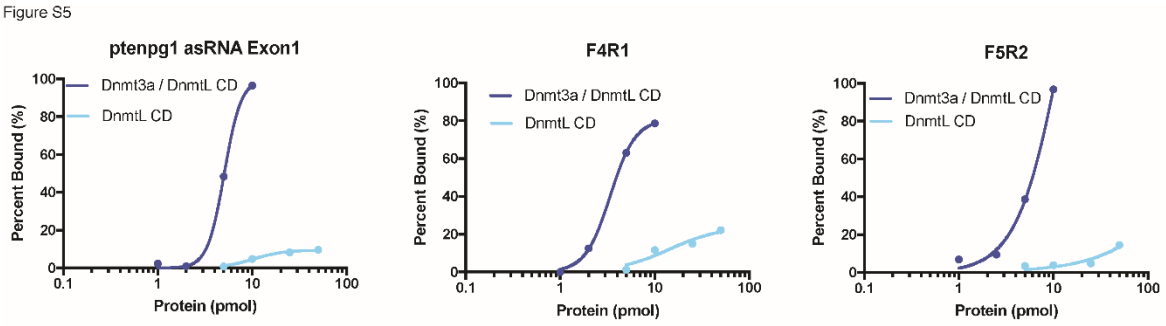


Figure S3-9 Analysis of the bound fraction of EMSA for *PTENpg1* asRNA exon 1, F4R1 and F5R2 binding to Dnmt3a in the presence of DnmtL-CD

Bands were quantified using ImageJ and plotted in a log scale with a curve fit using Prism 7.

Table S3-1 Reported EST BG772190 found on chr10:89623172-89623881

EST	Sequence
EST BG772190	5’TACCGCCCCCTGCCCTGCCCTGCCCTCCCCTCGCCCGGCGCGGTCCCG TCCGCCTCTCGCTCGCCTCCCGCCTCCCCTCGGTCTTCCGAGGCGCCCG GGCTCCCGGCGCGGCGGCGGAGGGGGCGGGCAGGCCGGCGGGCGGTG ATGTGGCGGGACTCTTTATGCGCTGCGGCAGGATACGCGCTCGGCGCT GGGACGCGACTGCGCTCAGTTCTCTCCTCTCGGAAGCTGCAGCCATGAT GGAAGTTTGAGAGTTGAGCCGCTGTGAGGCGAGGCCGGGCTCAGGCGA GGGAGATGAGAGACGGCGGCGGCCGCGGCCCGGAGCCCCTCTCAGCG CCTGTGAGCAGCCGCGGGGGCAGCGCCCTCGGGGAGCCGGCCGGCCTG CGGCGGCGGCAGCGGCGGCGTTTCTCGCCTCCTCTTCGTCTTTTCTAAC CGTGCAGCCTCTTCCTCGGCTTCTCCTGAAAGGGAAGGTGGAAGCCGT GGGCTCGGGCGGGAGCCGGCTGAGGCGCGGCGGCGGCGGCGGCACCT CCCGCTCCTGGAGCGGaGGGGAGAAGCGGCGGCGGCGGCGGCCGCGGC tGGCTGCAGCTCCAGGGAGGGGGTCTGAGTCGCCTGTCACCATTTCCAG GGCTGGGAACGCCGGAGAGTcGGTCTCTCCCCTTCTACTGnCTCCAACA CGGCCGGCGGCTGGCggcggcaACATCCAGGGACCCGGG3’

Table S3-2 Truncated *PTEN* alpha exon 1 variants

The experimentally determined 3a binding bulge is underlined in the F4R1 and full-length *PTENpg1* alpha exon 1 and the AT insertion in bold that disrupts the 3a binding bulge for the F5R2 variant.

Variant	Sequence
PTENpg1 **antisense alpha exon 1 (described in (Johnsson, Ackley, et al. 2013b))**	5’GGAAGAGGCTGCACAGTTAGAAAAGACGAAGAAGAAGCGAGA AACGCCGCCGCTGCCGGCGCCGCCCCCGCGGATGCTCACGGGTT GCTGAGAGGGGCTTCAGGCCGGGCCGGGCCGGGCCGCCGCCGCC GCCGCCGTCTCTGTCTCTCATCTCCCTCGCCTGAGCCCGGCCTCGC CTCACAGCGGCTCAACATTCAAACTTCCATCACGGCTGCAGCTTC TGAGAGGAGAGCATCTTCTAG3’
F2R1 PCR set 2 (variant 1 ***PTENpg1*** **asRNA alpha, 220bp fragment)**	5’CCGGCGCCGCCCCCGCGGATGCTCACGGGTTGCTGAGAGGGGC TTCAGGCCGGGCCGGGCCGGGCCGCCGCCGCCGCCGCCGTCTCT GTCTCTCATCTCCCTCGCCTGAGCCCGGCCTCGCCTCACAGCGGC TCAACATTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGGAGAG CATCTTCTAGTTTTT3’
F3R1 PCR set 3 (variant 2 ***PTENpg1*** **asRNA alpha, 164 bp fragment)**	5’GGGCCGGGCCGCCGCCGCCGCCGTCTCTGTCTCTCATCTCC CTCGCCTGAGCCCGGCCTCGCCTCACAGCGGCTCAACATTCAAAC TTCCATCACGGCTGCAGCTTCTGAGAGGAGAGCATCTTCTAGTTT TT3’

F4R1 PCR set 4 (variant 2 *PTENpg1* asRNA alpha, 107bp fragment)	5'GGGTAGTGAAGGCTAGCGGGCCTCGCCTCACAGCGGCTCAACA TTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGGAGAGCATCTT CTAGTTTTTGGTACCTATCTGA3'
F5R2 PCR set 4 (variant 2 *PTENpg1* asRNA alpha, 217bp fragment)	5'TAGTGAAGGCTAGCGCGCCGCCGCTGCCGGCGCCGCCCCCGCG GATGCTCACGGGTTGCTGAGAGGGGCTTCAGGCCGGGCCGGGCC GGGCCGCCGCCGCCGCCGCCGTCTCTGTCTCTCATCTCCCTCGCC TGAGCCCGGCCTCGCCTCACAGCGGCTCAACTCTCAAACTTCCAT C**AT**GGCTGCAGCTTCCGAGAGGTTTTTGGTACCTATCTGA3'

Table S3-3 The F4R1 functional domains

Several variations of the F4R1 transcript were derived. Deletion construct C1 contains a change to the major loop of F4R1, resulting in the replacement of the major loop with another loop sequence. Construct C2 contains a deletion of "lower" stem and construct C3 contains a deletion of the "upper" stem.

Variant	Sequence
F4R1	5'GGGTAGTGAAGGCTAGCGGGCCTCGCCTCACAGCGGCTCAACATTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGGAGAGCATCTTCTAGTTTTTGGTACCTATCTGA3'
C1 Construct	5'TAGTGAAGGCTAGCGGGCCTCGCCTCACAGCGGCTCACGAATGCAGCTTCTGAGAGGAGAGCATCTTCTAGTTTTTGGTACCTATCTGA3'
C2 Construct	5'CTCGCCTCACAGCGGCTCAACATTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGGAGAGC3'
C3 Construct	5'GGGTAGTGAAGGCTAGCGGGCGAATCTTCTAGTTTTTGGTACCTATCTGA3'

Table S3-4 Primers and ODN sequences used in analysis (5'-3' shown)

Primer	Sequence
PTENproF1	GCTGCAGCCATGATGGAAGTTTGA
PTENproR1	AAAGACGAAGAGGAGGCGAGAAAC
PTENproF2	TGATGTGGCGGGACTCTTTATGC
PTENproR2	TCACAGCGGCTCAACTCTCAAACT
PTENF3	AGAAAGCTTACAGTTGGGCCCTGT
PTENR3	GCCACAGCAAAGAATGGTGATGCT
PTENP1ex1_F	GGAAGAGGCTGCACAGTTA
PTENP1ex1_R	CTAGAAGATGCTCTCCTCTCA
paRNA_F	ATGTGGCGGGACTCTTTATG
paRNA_R	GCGGCTCAACTCTCAAACT
PTENex1_F	TGCCATCTCTCTCCTCCTT
PTENex1_R	CGAATCCATCCTCTTGATATCTCC
PTENex9_F	TGTAATCAAGGCCAGTGCTAAA
PTENex9_R	AGCATCCACAGCAGGTATTATG
PTEN unspliced F	AAAGCTGGAAAGGGACGAACTGGT
PTEN unspliced R	TCTCAGATCCAGGAAGAGGAAAGG
mirN367 (control) ODN	GTGTGGGGTTTTAGCTTCGTGAA
ODN1	GTCTCTCATCTCCCTCGCCT
ODN2	GCTTCCACCTTCCCTTTCAG
Scrambled ODN2	CTAACTCTCCGTGTCCTCCT

pcDNA3.1 F	CCCACTGCTTACTGGCTTATC
pcDNA3.1 R	CAGATGGCTGGCAACTAGAA
Killin F	ACACAAGCACCCACATCCAAA
Killin R	AGTCCTTTGGCTTGCTCTTAG

3.3.2 Materials and methods

3.3.2.1 Biotin ODN2 Immunoprecipitation

5' biotin-labelled ODN2 or scrambled control (**Table S3-4**) was transfected into HEK293 cells (as described in (Hawkins et al. 2009a; Han, Kim, and Morris 2007)). Cells were harvested 48 hours post-transfection and permeabilised in 0.25% Triton-X/PBS and then washed in PBS. Cells were treated with either RNase A, RNase H or no RNase control at 37°C for 15 minutes. RNase inhibitors were added and cells were cross-linked with formaldehyde. Cells were resuspended in lysis buffer (1% SDS, 50mM Tris-HCl (pH 8), 10mM EDTA, RNase inhibitors) and sonicated. Immunoprecipitation was performed using Dynabeads™ MyOne™ Streptavidin Beads (Thermo Fisher). Beads were incubated with the samples for 30 minutes and then washed five times in lysis buffer. Samples were eluted at 95°C for 15 minutes in elution buffer (1% SDS, 10mM EDTA, 0.1mM $NAHCO_3$, pH 8). Samples were then divided and treated with either RNase or DNase. Enrichment of DNA in RNase-treated samples was determined by qPCR using the following primer sets: PTENP1_F/R, paRNA_F/R, PTENex1_F/R, PTENex9_F/R (**Table S3-4**). The RNA in DNase treated samples was quantified by RT-qPCR using the same primer sets as above. Scramble control was subtracted from the IP and data was normalised to the sample input.

3.3.2.2 *PTENpg1* asRNA Exon1 and truncated variant transcription

PTENpg1 asRNA exon one and truncated variants (**Table S3-2**) were *in vitro* transcribed from linearised pcDNA 3.1 plasmids (containing the various inserts) using T7 RNA polymerase (Batey, Sagar, and Doudna 2001a; Batey, Sagar, and Doudna 2001b). RNA constructs were purified from transcription components by denaturing gel electrophoresis.

3.3.2.3 Electrophoretic Mobility Shift Assays

Protein lncRNA binding reactions were performed in a final volume of 35μL and contained 637, 500, 700, or 600 ng *PTENpg1* asRNA exon1, F4R1, F5R2, or GFP RNA, respectively. All lncRNA variants were refolded by heating at 90 °C for 80 s before snap cooling on ice. Refolding was performed in a buffer containing 50 mM Hepes at pH 7.5, 150 mM KCl, 1.5 mM MgCl2, 1 mM TCEP, and 20% glycerol for all lncRNAs. After refolding, protein was incubated with lncRNA for 30 min on ice. Dnmt3a/DnmtL-CD binding reactions contained between 1× and 10× molecular excess of the protein to RNA. DnmtL-CD binding reactions between 5× and 50× the molecular excess of the protein to RNA. Samples were loaded onto a 4.5% TBE acrylamide gel (containing 2.5 mM MgCl2) and run at 250 V for 4 h at 4 °C. Gels were stained with SYBR gold (Invitrogen) and visualized using a Typhoon FLA 900 biomolecular imager.

3.3.2.4 DNMT3a/DNMTL-CD Purification

Dnmt3a (residues 284–910) and DnmtL-CD (residues 178–379) were cloned into a modified petDuet vector, which is designed for the coexpression of two target genes. Dnmt3a/DnmtL-CD proteins were coexpressed in Escherichia coli strain Rosetta 2 (DE3). The transformants were grown at 37 °C in LB medium and induced at an OD600 of 0.6 with IPTG and further incubated for 20–24 h at 18 °C. Dnmt3a/DnmtL-CD was purified from the supernatant of the cell lysate by three-step liquid chromatography. Nickel affinity, heparin affinity, and gel filtration chromatography were used and the purified protein complex stored in 50 mM Hepes at pH 7.5, 150 mM NaCl, 10% glycerol, and 1 mM TCEP. The purified proteins were estimated to be >90% pure by Coomassie bluestained SDS/PAGE and were concentrated to 2–3 mg/mL for electrophoretic mobility shift assays. DnmtL-CD (residues 178– 379) was cloned into pLIC-HK vector, which is for single gene expression, and overexpressed using

the same methods as Dnmt3a/DnmtL-CD. DnmtL-CD was purified from cell lysate using Ni2+ affinity and gel filtration chromatography and stored in 50 mM Hepes at pH 7.5, 150 mM NaCl, 10% glycerol, and 1 mM TCEP. Purified DnmtL-CD was estimated to be >95% pure and was concentrated to 9–10 mg/mL.

3.3.2.5 ODN Transfections

Control (mirN367), ODN1 and ODN2 were transfected into cells to a final concentration of 100 nM using Lipofectamine 2000 (Life Technologies, Carlsbad, CA, USA). ODNs were transfected into both HEK293 and the *PTENpg1* asRNA exon1 over-expressing 293-HEK PG1 cells.

3.3.2.6 qRT-PCR

Quantitative Real Time PCR (qRT-PCR) analysis was performed using KAPA SYBR FAST qPCR Master Mix (KAPA Biosystems, Woburn, MA, USA). Plate was placed in ViiA 7 Real-Time PCR system (Life Technologies, Carlsbad, CA, USA). Cycling conditions: 95°C for 2 mins, then 40 cycles of 95°C for 3s and 60°C for 30s.

3.3.2.7 Directional RT analysis of gene expression

Directional reverse transcription (directional RT) was performed on HeLa total RNA using the PTENproR1 primer (**Table S3-4**) to strand specifically convert any RNA transcripts originating from the *PTEN* 5' UTR into ssDNA. The resulting product was PCR amplified using primer set PTENproF2/R2 and run on a 2% agarose gel (**Figures 3-1B**).

3.3.2.8 Chromatin Immunoprecipitation

ChIP analysis was carried out in HEK293 cells using anti-DNMT3a (Abcam, Cambridge, MA, USA) (Abcam, cat. no. ab2850). The CHIP was performed 48 hours post-transfection with ODN's (**Figures 3-2D - I**) or various biotin labelled truncations of the full-length

PTENpg1 alpha exon 1 transcript (**Figure 3-3A**) following previously described techniques (Weinberg et al. 2006; Saayman et al. 2014b; Lister, Clemson, and Morris 2015). The relative enrichment of DNMT3a was determined at the *PTEN* promoter using primer sets PTENproF2/R2 (**Figure 3-1F, Table S3-4**) or PTENproF1/R1 (**Figure 3-3B and Table S3-4**). Any IgG or no antibody values are first subtracted from the resultant IP and input values and then each sample is standardized relative to the sample input.

3.3.2.9 T7-transcribed synthetic RNA pulldown in presence of ODNs

Synthetic biotinylated ncRNAs were generated by T7 transcription using the Ampliscribe™ T7-Flash™ Biotin –RNA Transcription Kit (Epicentre® Biotechnologies, WI, USA) according to the manufacturer's instruction. Templates for T7 transcription were prepared by PCR of pcDNA3.1 plasmids expressing the relevant ncRNAs. Primers pcDNA3.1 F/R were used for PCR amplification (**Table S4**). *PTENpg1* alpha exon 1 was *in vitro* transcribed using Durascribe with biotin linked dCTPs (described in detail in (Saayman et al. 2014b; Saayman et al. 2014a)). The resultant biotin *PTENpg1* α exon one transcript was transfected into HEK293 cells at a concentration of 100nM 18 hours after transfection of either control ODN mirN367 or ODN2 transfection (see ODN Transfections above). 30 hours post-transfection of biotin conjugated transcripts, cells were cross-linked with formaldehyde at 1% final concentration for 10 minutes at room temperature followed by addition of glycine to a final concentration of 0.125M and a further incubation for 10 minutes at room temperature. Cells were then lysed with ChIP lysis buffer (5mM PIPES, 85mM KCl and 0.5% NP40) supplemented with PMSF on ice for 20 minutes. Chromatin was sheared by sonication. Cell lysates containing sheared chromatin were incubated with Dynabeads® M280™ Streptavidin (Life Technologies, CA, USA) prepared according to the manufacturer's instructions for 1 hour on a rotating platform. Beads were pulled down with a magnet for 3 minutes and washed with Low salt immune complex wash buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA;

20 mM Tris-HCl, pH 8.1; 150 mM NaCl); high salt immune complex wash buffer (0.1% SDS; 1% Triton X-100; 2 mM EDTA; 20 mM Tris-HCl, pH 8.1; 500 mM NaCl); LiCl Immune complex wash buffer (0.25 M LiCl; 1% NP40; 1% sodium deoxycholate; 1 mM EDTA; 10 mM Tris-HCl, pH 8.1); and TE buffer (10 mM Tris-HCl; 1 mM EDTA, pH 8.0). Each wash step was carried out for 3 minutes on a rotating platform. Streptavidin bead-biotinylated RNA-DNA complexes were re-suspended in Elution buffer (1% SDS, 0.1M $NaHCO_3$) and heated at 95°C for 5 minutes to dissociate biotin RNA from beads. Eluted biotinylated RNA complexes were removed from beads via magnet pull down and supernatants were analysed by qPCR for enrichment at the *PTEN* promoter. Samples were then DNase, RNase A and Proteinase K treated so as to remove everything from elutes except dsRNA's. These were reverse transcribed and analyzed by qPCR for *PTEN* promoter enrichment. Primer sets used for analysing the *PTEN* promoter were PTENproF1/R1, PTENproF2/R2 and PTENproF2/R1 (**Table S3-4 and Figure 3-2C**).

3.3.2.10 Truncated PTEN alpha exon 1 and F4R1 deletions CHIP

Various truncated versions of *PTEN* alpha exon 1 (**Table S3-2**) and mutants of F4R1 (**Table S3-3**) were generated to be expressed from the CMV promoter in the context of pcDNA3.1 (Genewiz, Carlsbad CA USA). The constructs (**Tables S3-2 and S3-3**) were *in vitro* transcribed using Durascribe with biotin linked dCTPs (described in detail in (Saayman et al. 2014b; Saayman et al. 2014a)). The resultant biotin-labelled transcripts and a GFP-biotin control were transfected into HEK293 cells (50nM) and assessed by RIP 48 hours later using PTENproF2/R2 (**Table S3-4**) primers for detection at the *PTEN* promoter.

3.3.2.11 *PTEN* promoter CpG methylation post-ODN2 treatment

Genomic DNA was isolated from Hek293 cells after 48h treatment with ODN scrambled control and ODN2 (**Table S3-4**). Briefly, 200 ng of DNA was digested with the methyl-

cytosine specific restriction enzyme McrBc (New England Biolabs) overnight at 37°C. Twenty-four hours later the enzyme was heat inactivated at 65°C for 1h and qRT-PCR performed (Kapa Biosystems using Killin F/R primers (**Table S3-4**) and standardized to uncut input. The delta CT values were converted to fold-change values and the ratio between ODN2/ODN scrambled control treated cells was calculated.

4 RNA directed SNPs in human cells

Chapter contents:

Page No.

List of tables:

Page No.

List of figures:

Page No.

4.1 RNA directed generation of single nucleotide polymorphisms in human cells

Authors:

Nicholas Lister, Matthew Clemson, Denis O'Meally and Kevin V. Morris

Current status of paper:

In preparation

Date paper accepted for publication or anticipated date of publication:

August 2018

Extent to which research is your own:

I performed ChIP experiments for DUT, APOBEC3A, TDG, MBD4, methyl-CpG, SMUG, UNG and DNA Polymerase, as well as experiments using shRNAs for TDG, MBD4 and SMUG. I carried out expression profiling (qPCR) and western blots for cells with and without MBD4 and SMUG knockdown. I performed experiments looking for the generation of SNPs following the transfection/transduction of an shRNA or lncRNA into MBD4 or SMUG KD cells. Bioinformatics analysis were conducted by other people. I performed all statistical tests and data analysis for these experiments.

Your contribution to writing the paper:

I was involved in the collaborative writing of this draft of the manuscript. I also prepared a number of figures in the draft manuscript, see figures 2a-c and all of figure 3.

If paper has not yet been accepted, has the paper been rejected by any journals:

No

Comments:

This manuscript presents data demonstrating that ncRNAs are capable of directing mutations from cytosine to thymine around target sites (*PTEN* and *UBC* gene promoters) in human cells when proteins involved in the deamination repair pathway, such as MBD4 and SMUG, are repressed. This disturbs other processes, such as preventing transcription factor binding, in

the cell. Heritable mutations such as these could be passed down through generations and may be responsible for driving sequence evolution, suggesting a larger role for RNA in SNP genesis than previously known. This manuscript is still in the drafting stage, as there are more SNP deep sequencing experiments being performed for validation (see **Section 5.6**).

4.2 Manuscript draft

RNA directed generation of single nucleotide polymorphisms in human cells

Authors: Nicholas Lister[1], Matthew Clemson[1], Denis O'Meally[2], and Kevin V. Morris[*,2]

*Corresponding author:

[1]The University of New South Wales, Biotechnology and Biomedical Sciences, Sydney NSW 2052 Australia

[2]Center for Gene Therapy, City of Hope – Beckman Research Institute and Hematological Malignancy and Stem Cell Transplantation Institute at the City of Hope. 1500 E. Duarte Rd., Duarte, CA, 91010, USA (kmorris@coh.org).

Key words: non-coding RNA, selection, DNA methylation, deamination, SMUG, MBD4.

4.2.1 Abstract

Experimental observations over the last decade have ascribed distinct functions to non-coding RNAs. While there are various thematic roles for these transcripts, the emerging notion is that they play a unique role in modulating epigenetic state and regulating transcription. In human cells, non-coding RNAs have been found to direct epigenetic complexes (including some capable of inducing DNA methylation) to target loci. Interestingly, methylated cytosine's undergo the process of deamination to remove this methylation, which if not properly repaired by various proteins becomes recognized to the cell as a thymine. We find here that non-coding RNAs can direct cytosine to thymine mutations at particular loci in human cells when MBD4 or SMUG, enzymes involved in the deamination repair pathway, are repressed. Collectively, the observations presented here suggest that RNA plays an underappreciated role in the emergence of single nucleotide polymorphism genesis in human cells, and by extension the cytosine to thymine content of the genome.

4.2.2 Introduction

The collective viewpoint as to the complexities and nuances involved in the functional aspects of regulating the cell has been radically altered as a result of the exciting findings generated from the Fantom (Carninci et al. 2005) and Encode consortiums (Pennisi 2012; Rosenbloom et al. 2012). These consortiums demonstrated that at least 60-70% of the human genome is transcribed into non-coding transcripts. Many of these non-coding transcripts appear functional in the cell (Morris and Mattick 2014a, 2014b). Studies carried out in human cells have determined that both small (<200 bp)and long (>200 bp) classes of non-coding RNAs can interact with proteins involved in epigenetic processes that regulate chromatin states, gene accessibility, and transcription (**Figures 4-1A-B**) (Johnsson et al. 2014; Morris and Mattick 2014a; Johnsson, Lipovich, et al. 2013).

Experimental observations suggest that both long (Johnsson, Ackley, et al. 2013b; Johnsson et al. 2014; Johnsson, Ackley, et al. 2013a) and small non-coding RNAs (Weinberg et al. 2006) can functionally recruit DNMT3a (DNA methyltransferase 3a) to target loci. Small non-coding RNAs were shown to guide DNA methylation and stable epigenetic silencing of homologous-targeted loci *in vitro* (Hawkins et al. 2009a). DNA methylation occurred at the shRNA targeted promoter for Ubiquitin ligase (*UBC*) after ~5 days of sustained shRNA targeting (Hawkins et al. 2009a, 2009b) (**Figures 4-1A-C**).

Under normal circumstances, the shRNA directed cytosine methylation would be subjected to a removal of the methyl group on the cytosine and deamination base repair (**Figures 4-1C-H**). However, if this process is blocked or disrupted, some of the methylated cytosine's may undergo a hydrolysis reaction resulting in the production of ammonia, and the conversion of the methylated cytosine to a uracil in the DNA sequence (**Figure 4-1I**). Following DNA replication, the cell will recognize this uracil in the DNA as a thymine (T), which will become fixed in one of the two daughter strands (Lutsenko and Bhagwat 1999;

Shen, Rideout, and Jones 1994) (**Figures 4-1I-K**).

While a cytosine to thymine single nucleotide polymorphism (SNP), is considered a random event, the spontaneous deamination of methylated cytosine's in the human genome has been found to be ~ 2 fold faster than non-methylated cytosine's (Shen, Rideout, and Jones 1994), suggesting that there is a bias toward cytosine methylation at CpG's in the deamination process. Moreover, the most frequent mutation in human cancers appears to be the emergence of a C to T conversion (Neddermann et al. 1996; Alexandrov et al. 2013). These conversions may be related to APOBEC3A (Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3A) function (Alexandrov et al. 2013), a protein known to be involved in deamination of DNA (Wijesinghe and Bhagwat 2012).

We hypothesise here that the emergence of C to T conversion in the human genome may be the result of RNA directed DNA methylation being incorrectly repaired due to a reduction or loss of those proteins involved in the deamination pathway (Morris 2015). We find here transfecting/transducing small hairpin (shRNA) and long-non-coding RNAs (lncRNA) into human cell lines with loss of MBD4 and SMUG (two proteins involved in base excision repair of methylated cytosine via deamination) (**Figure 4-1 I-K**) results in the emergence of a SNP, that is ~200 bp distal of the non-coding RNA target loci.

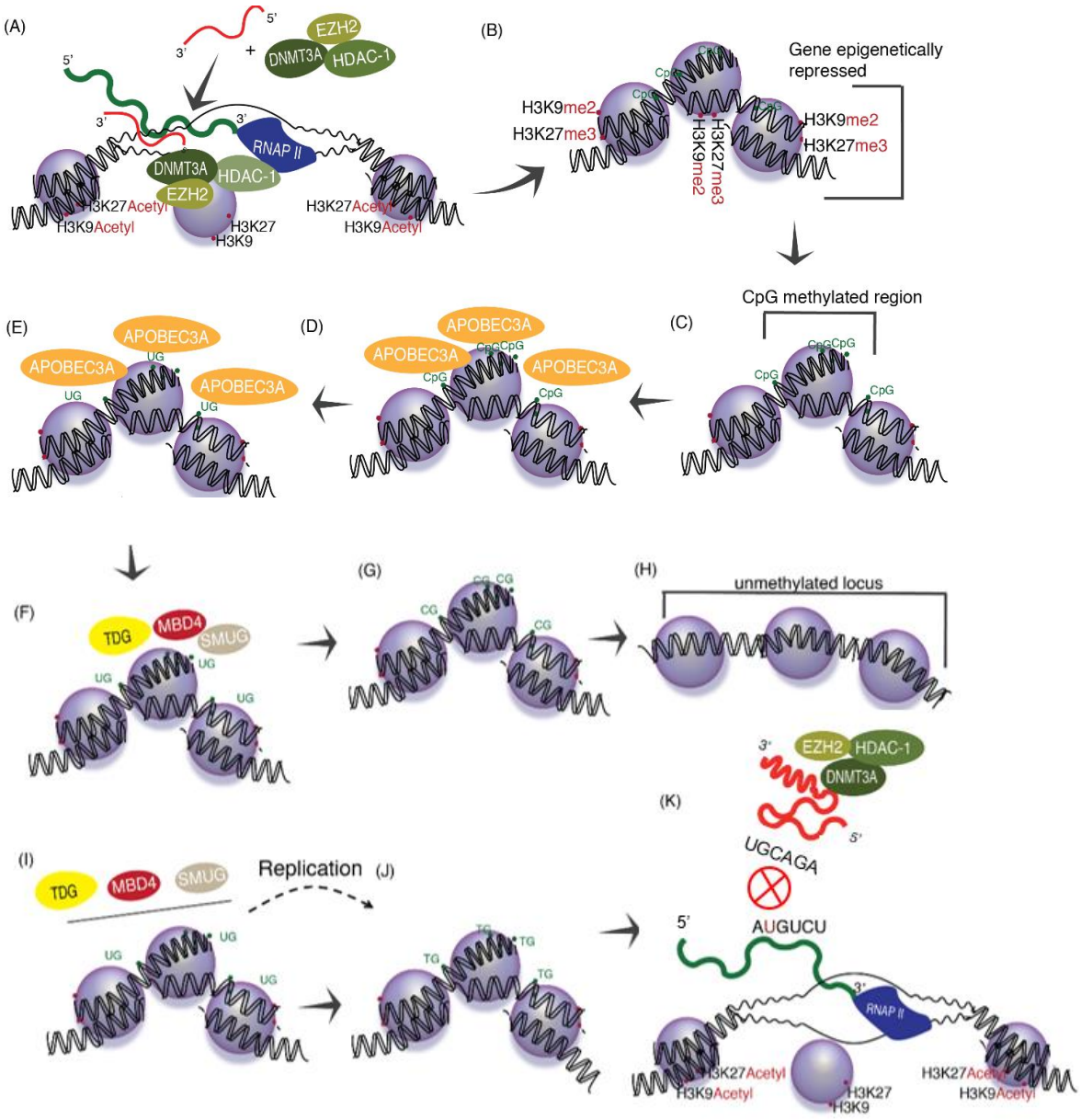

(A)
5'
3'
EZH2
DNMT3A
HDAC-1
RNAP II
H3K27Acetyl
H3K9Acetyl
H3K27
H3K9
(B)
Gene epigenetically repressed
H3K9me2
H3K27me3
(C)
CpG methylated region
CpG
(D)
APOBEC3A
(E)
UG
(F)
TDG
MBD4
SMUG
(G)
CG
(H)
unmethylated locus
(I)
Replication
(J)
TG
(K)
UGCAGA
AUGUCU

Figure 4-1 Mechanism of RNA directed gene mutation

The genesis of SNPs in the human genome is hypothesized to function through RNA guided CpG methylation and faulty deamination repair as shown schematically. (**A**) A *de novo* derived small or long non-coding antisense RNA with homologous targeting to a particular transcribed locus can recruit and/or interact with various proteins involved in epigenetic silencing, e.g. DNMT3a, EZH2 and HDAC-1. (**B**) This targeting results in epigenetic silencing of the targeted locus including histone methylation H3K9me2 and H3K27me3 as well as (**C**) CpG DNA methylation. (**D**) The methylated cytosines at or around the targeted RNA undergo deamination by the actions of APOBEC3A, whereby (**E**) the methylated cytosine is converted to a uracil. (**F-H**) The resulting uracil/guanine pairing mistake is repaired by TDG, MBD4 and SMUG, but if (**I**) this process is disrupted and/or replication occurs before repair of the locus is completed then a thymine residue can become fixed in the DNA such that (**J**) a permanent cytosine to thymine change has occurred in the genome which could alter downstream RNA/RNA interactions (shown) or DNA/protein or RNA/protein interactions.

4.2.3 Results

To test if reduced expression of proteins involved in the repair of deaminated 5mC (methylated cytosine) results in the of C to T mutation at non-coding RNA over-expressed loci, we capitalized on an inducible shRNA system for the ubiquitin ligase gene (Hawkins et al. 2009a, 2009b). In this system, the addition of Tetracycline (Tet) induces shRNA UBC 167 (sh167) expression, which leads to CpG methylation of the *UBC* promoter and transcriptional gene silencing (Hawkins et al. 2009a, 2009b). To determine the extent that sh167 directed CpG methylation results in the recruitment of proteins involved in base excision repair, shRNA 167 was induced and an enrichment of various proteins determined by chromatin immunoprecipitation (ChIP) qPCR.

We observed an increase in DUTP pyrophosphatase (DUT), APOBEC3A, MBD4, SMUG and uracil DNA glycosylase (UNG) (**Figures 4-2A-C**) at the *UBC* promoter. However, consistent with previous observation in these cells (Hawkins et al. 2009a, 2009b), increased sh167 expression lead to increased CpG methylation at the sh167 target site (**Figure 4-2B**). These data suggest that sh167 targeting of the *UBC* promoter results in the enrichment of CpG methylation as well as key proteins involved in the 5mC deamination repair pathway. Experimental variation was higher than expected here for MBD4, SMUG1 and UNG, and could be resolved with further experimental replicates to improve accuracy of the measurements.

To determine the extent the loss of these key deamination repair proteins causes C to T conversion, RNAi was used to repress the expression of MBD4, SMUG and TDG (**Figure 4-2D**). There was higher than expected variation observed here for MBD4, SMUG and TDG, which would also improve in accuracy with further experimental replicates. A new C to T conversion emerged when these cultures (with sh167 expression induced) were treated with RNAi to MBD4 and SMUG1 for 1 week. This new SNP (called UBC22), emerged in the

sh167 treated cells, but not the control cells or in those cells with repressed TDG expression. UBC22 was ~200bp downstream of the sh167 target site (**Figure S4-4**) and was coincidentally within an SP1 transcription factor binding site (**Figure 4-2E**). To determine the affect UBC22 had on SP1 binding, the control vs sh167 were assessed by ChIP with an SP1 antibody following RNAi to MBD4. The UBC22 SNP resulted in an 87% loss of SP1 binding at the UBC22 site (**Figure 4-2F**). Collectively, these data suggest that small non-coding RNAs have the ability, when targeted to genomic loci in the absence of MBD4 and SMUG1, to modulate the emergence of a C to T SNP, which has functional significance.

Figure 2

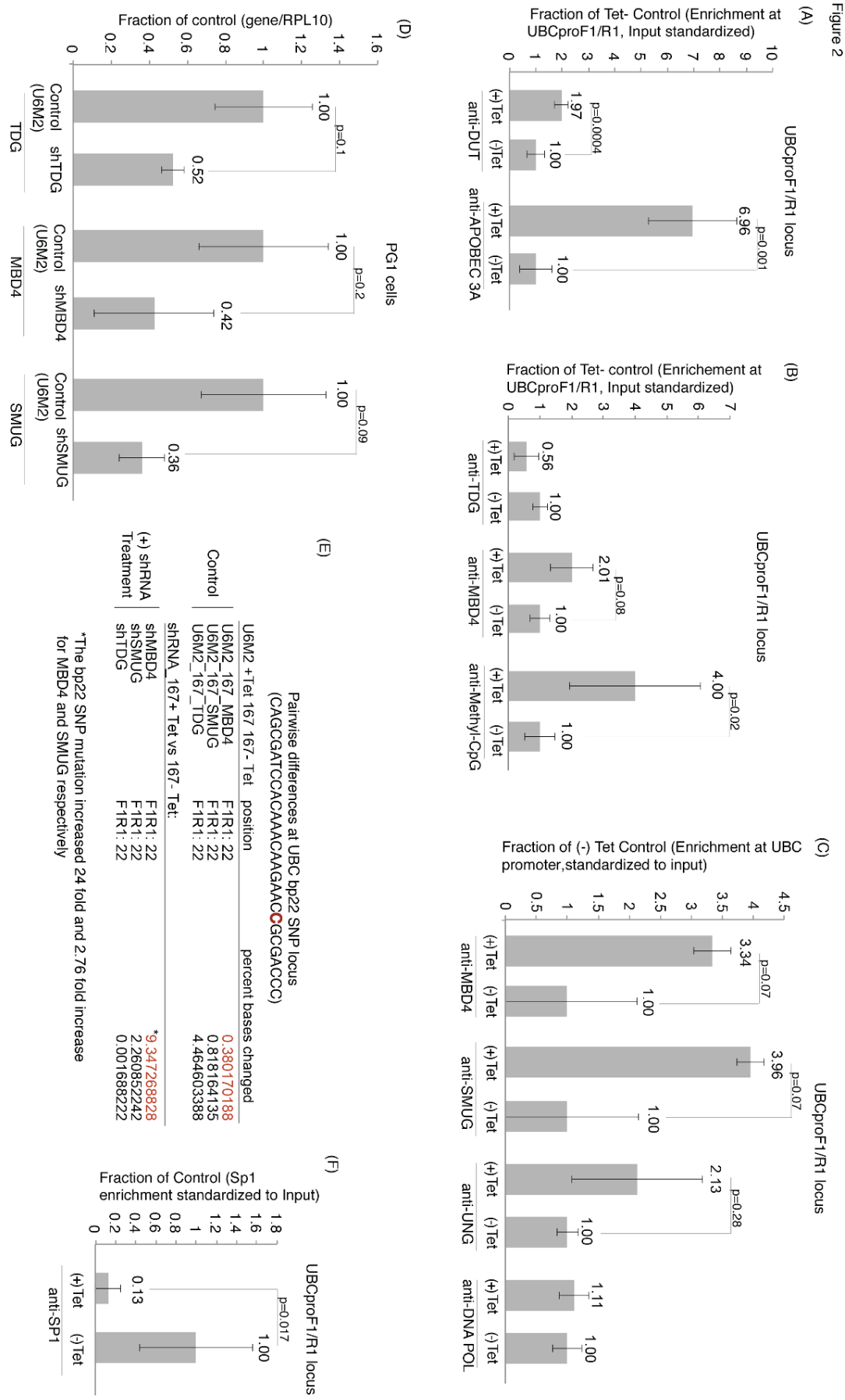

(A)
Fraction of Tet- Control (Enrichment at UBCproF1/R1, Input standardized)
UBCproF1/R1 locus
anti-DUT
anti-APOBEC 3A
(+) Tet
(-)Tet
1.97
1.00
6.96
1.00
p=0.0004
p=0.001
(B)
Fraction of Tet- control (Enrichement at UBCproF1/R1, Input standardized)
UBCproF1/R1 locus
anti-TDG
anti-MBD4
anti-Methyl-CpG
0.56
1.00
2.01
1.00
4.00
1.00
p=0.08
p=0.02
(C)
Fraction of (-) Tet Control (Enrichment at UBC promoter,standardized to input)
UBCproF1/R1 locus
anti-MBD4
anti-SMUG
anti-UNG
anti-DNA POL
3.34
1.00
3.96
1.00
2.13
1.00
1.11
1.00
p=0.07
p=0.07
p=0.28
(D)
Fraction of control (gene/RPL10)
PG1 cells
TDG
MBD4
SMUG
Control (U6M2)
shTDG
Control (U6M2)
shMBD4
Control (U6M2)
shSMUG
1.00
0.52
1.00
0.42
1.00
0.36
p=0.1
p=0.2
p=0.09
(E)
Pairwise differences at UBC bp22 SNP locus
(CAGCGATCCACAAACAAGAACCGCGACCC)
U6M2 +Tet 167 167- Tet position percent bases changed
U6M2_167_MBD4 F1R1: 22 0.380170188
U6M2_167_SMUG F1R1: 22 0.818164135
U6M2_167_TDG F1R1: 22 4.464603388
shRNA_167+ Tet vs 167- Tet:
(+) shRNA Treatment
shMBD4 F1R1: 22 *9.347268828
shSMUG F1R1: 22 2.260852242
shTDG F1R1: 22 0.001688222
Control
*The bp22 SNP mutation increased 24 fold and 2.76 fold increase for MBD4 and SMUG respectively
(F)
Fraction of Control (Sp1 enrichment standardized to Input)
UBCproF1/R1 locus
anti-SP1
(+) Tet
(-) Tet
0.13
1.00
p=0.017

Figure 4-2 Induction of SNPs by sh167 targeted silencing of *UBC*

(**A-C**) ChIP analysis of the *UBC* promoter 72 hours post-sh167 induction. Enrichment of (**A**) DUT, APOBEC3A, (**B**) TDG, MBD4, methyl CpG, and (**C**) MBD4, SMUG1, UNG and DNA polymerase at the UBC 167 shRNA target locus in the *UBC* promoter are shown as determined by ChIP. (**D**) Small hairpin RNA silencing of TDG, MBD4 and SMUG1. (**E**) Deep sequencing analysis of the UBC sh167 target site at day 7 post-Tet induction of sh167 and shRNA knockdown of MBD4, TDG, and SMUG1. (**F**) SP1 enrichment at the UBC SNP 22 site as determined by ChIP following MBD4 small hairpin silencing. For A-D and F the averages of triplicate treated cultures are shown with the standard deviations and p values from a paired T-test.

The ability of small RNAs to induce 5mC raised the question of if endogenous long non-coding RNAs can also modulate 5mC deposition, and subsequently the emergence of C to T conversion. We examined the *PTEN* pseudogene 1 α antisense (*PTENpg1*) transcript (242 bp), a known modulator of *PTEN* (Johnsson, Ackley, et al. 2013b, 2013a). Direct RNA:RNA interactions between a *PTEN* promoter associated 5' UTR transcript and *PTENpg1* α lead to epigenetic regulation of *PTEN* (Lister et al. 2017). We studied the ability of this transcript to induce SNP genesis at the *PTEN* promoter, in comparison with sh167 targeting of the *UBC* promoter.

To examine the extent that endogenous ncRNAs cause C to T conversion in human cells undergoing faulty DNA methylation repair pathways, MBD4 or SMUG1 knockout HAP1 cells were utilized. These cells have undergone a homozygous CRISPR mutation (2 bp deletion in coding exon) in the MBD4 or SMUG1 genes, and were verified by western blot to exhibit MBD4 or SMUG1 inhibition (**Figure 4-3A**). We also generated HEK293 MBD4 knockout cell lines to contrast the HAP1 cells. These cells exhibited a 10 to 100-fold loss of MBD4 (measured by qRT-PCR - **Figure 4-3B**), and protein inhibition also observed on a western blot (**Figure 4-3A**). The repression of MBD4 did not affect the ability of three non-coding RNAs (sh167, *PTENpg1* α and as6, a short antisense RNA (Lister, Clemson, and Morris 2015) to induce transcriptional gene silencing of their targeted loci (**Figure 4-3C**).

To determine if either a lncRNA (*PTENpg1* α) or shRNA (shUBC167) could induce a SNP at their target promoter (*PTEN* or *UBC*), HAP1 cells were treated with control, lncRNA or shRNA expressing plasmids (Treatment 1, see **Section 4.3.1.8**). After 1 week of treatment the cells were assessed for SNPs. Notably, a SNP was observed at the *PTENpg1* asRNA exon 1 target promoter in treated cultures relative to controls (across barcoded oligos PTEN5/6, **see Table 4-1 and 4-2**) that was interestingly ~200bp downstream of the lncRNA target locus in the *PTEN* promoter (**Figure S4-5**). A SNP was also observed (found on the

reverse strand, hence G -> A mutation) in the shUBC167 treated cultures relative to controls (across barcoded oligos UBC3/4, **see Table 4-1**). Interestingly, this was also ~200bp away from the shRNA target locus in the *UBC* promoter (**Figure S4-6**). An average ~2.11-fold increase in C to T conversion was observed for the *PTENpg1* α lncRNA in both MBD4 and SMUG1 knockout cells, whereas for the sh167 shRNA an average ~330-fold increase was observed between control and pHIV7-IMPDH-U6-sh167 in SMUG1 knockout cells. The sh167 SNP observed here was not the same SNP in **Figure 4-2E**.

In **Table 4-1**, *PTENpg1* asRNA exon 1 directed SNPs at the *PTEN* promoter, and shUBC167 directed SNPs at the *UBC* promoter, in ΔMBD4 and ΔSMUG1 HAP1 cells. Reference gene refers to the barcoded oligos PCR product, while position indicates the nucleotide in the PCR product where the SNP occurred. Frequency refers to how often the altered base is found in the PCR product. Control is pHIV7-IMPDH-U6-mock. These data suggest that both lncRNAs and shRNAs are capable of directing C to T conversion in the absence of MBD4 or SMUG, though shRNAs appear to be more potent.

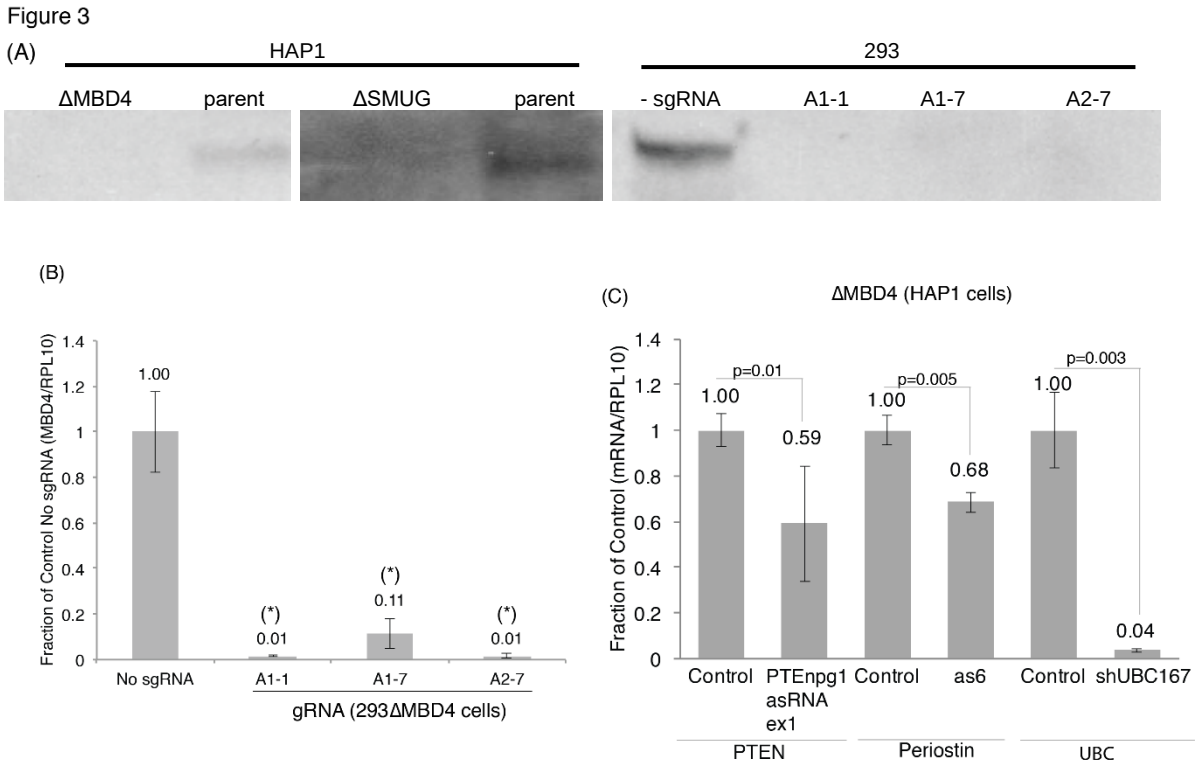


Figure 4-3 ShRNA, antisense lncRNA, and small antisense RNA targeting in MBD4 knockout cell lines

(**A**) MBD4 protein expression is stably repressed in HAP1 ΔMBD4, ΔSMUG1 and HEK293 ΔMBD4 cells as determined by western blot analysis. (**B**) MBD4 expression is lost in HEK293 ΔMBD4 cells. The A1-1, A1-7 and A2-7 clonal HEK293 ΔMBD4 cells were assessed relative to control, no gRNA control cells. (**C**) *PTENpg1* (Johnsson, Ackley, et al. 2013b), Periostin as6 (Lister, Clemson, and Morris 2015) and UBC-167 (Hawkins et al. 2009a) repress their target genes in ΔMBD4 HAP1 cells. Triplicate transfected ΔHAP1 cells were assessed for *PTEN*, *Periostin*, and *UBC* mRNA expression by qRT-PCR. For B-C the averages of triplicated treated HAP1 cells are shown with the standard deviations and p values from a paired T-test. The (*) indicates significance $p<0.05$.

Table 4-1 Induction of SNPs by lncRNA and shRNAs in MBD4 and SMUG knockout cell lines

Cell line	Ref Gene	Transient vs Stable	Position	Reference Base	Altered Base	Control (Frequency)	Treatment (Frequency)	Fold increase in SNP mutation	P Value (Fisher's exact test, one-tailed)
ΔMBD4	PTEN5	Transient	79	C	T	58.16%	99.95%	1.71	0.0
ΔMBD4	PTEN6	Transient	40	C	T	6.31%	23.84%	3.77	4.35E-185
ΔSMUG	PTEN5	Transient	79	C	T	55.27%	99.67%	1.80	0.0
ΔSMUG	PTEN6	Stable	40	C	T	16.96%	20.12%	1.18	0.0066
ΔSMUG	UBC3	Stable	105	G	A	0.08%	5.92%	74	2.98E-35
ΔSMUG	UBC4	Stable	97	G	A	0.01%	5.85%	585	1.20E-142

Table 4-2 Identified SNP and flanking sequence (around the C to T conversion) in the barcoded oligonucleotides from Table 1. SNP is shown in red.

Ref Gene	C to T conversion and flanking sequence
PTEN5/6	GCCGTGGGCT**T**GGGCGGGAGC
UBC3/4	CGATCACAGC**A**ATCCACAAAC

To determine if any endogenous C to T conversions occur in MBD4 or SMUG1 knockout cells (HAP1 and HEK293) and if they are present at known ncRNA target sites throughout the genome, cells were synchronised and genomic DNA sent for whole exome sequencing (Treatment 2, see **Section 4.3.1.9**). Bioinformatic analysis using both Varscan and MuTect 2 were used to overlay data sets, showing any common SNPs found between different knockout cell lines (**Figure S4-7**). Multiple intersects were run on Mutect 2 data using IGV (Integrative Genomics Viewer) and the UCSC (University of California, Santa Cruz) genome browser to determine shared SNPs between data sets (ΔMBD4 HAP1 vs ΔSMUG1 HAP1; ΔMBD4 HAP1 vs ΔMBD4 HEK293). Forty-three SNP sites were manually annotated, with almost all found in the 3' UTR of genes. These 3' UTR located SNPs were always 5' of a microsatellite sequence, which contained conserved repetitive nucleotides (**Figure S4-8**). 33% of these microsatellite expansion sequences consisted of all A and C nucleotides, 21% were either all A and T or all A and G nucleotides, and 18% were all T and G nucleotides. Every SNP site also contained ESTs derived from both the sense and antisense strands of the 3' UTR (**Figure S4-9)**. These data suggest that C to T conversions play a role in generating microsatellite indels via expansions and deletions.

4.2.4 Discussion

The findings here suggest that ncRNAs can mediate the generation of SNPs if the 5mC deamination repair pathway is disrupted. Previous studies have shown ncRNAs direct epigenetic regulation of the genome by recruiting chromatin/DNA modifying proteins to a target locus, altering chromatin states (i.e. deposition of 5mC) and transcriptional expression of a protein-coding gene (Hawkins et al. 2009a; Johnsson, Ackley, et al. 2013b). This represents the first step in the proposed process that leads to SNP generation within the genome. Subsequent 5mC deamination occurs so as to restore the original epigenetic state of the genome (Wijesinghe and Bhagwat 2012; Carpenter et al. 2012; Hashimoto, Hong, et al. 2012; Jacobs and Schar 2012). Both DNMT3a and HDAC-1 play a role in these two processes (Morris 2015), and MBD4 expression is tightly linked to the family of DNMTs (DNA methyltransferases, including DNMT3a) expression in human germ cells, suggesting MBD4 may control or maintain DNMT3a methylation (Galetzka et al. 2007).

The data presented here shows higher C to T mutations in the absence of MBD4 and SMUG1 proteins, particularly in the presence of an shRNA. MBD4 defects have been observed previously to affect the mutational landscape, and perturbing its function may contribute to tumorigenesis (Tricarico et al. 2015). Expanding upon their known function to epigenetically regulate gene expression (Hawkins et al. 2009a; Johnsson, Ackley, et al. 2013b; Morris and Mattick 2014a), ncRNAs were observed to be capable of directing the generation of these C to T SNPs. Sh167 (observed via ChIP) directed the known deamination proteins DUT, APOBEC3A, MBD4 and SMUG1 to the *UBC* promoter, as well as increased CpG methylation at the same locus (**Figures 4-2A-C**).

When cells underwent RNAi for MBD4 or SMUG1 in the presence of sh167, a C to T mutation was observed roughly 200bp downstream of the sh167 target site (**Figure 4-2E, S4-4**). Interestingly, this SNP was within a transcription factor binding site (SP1), which

subsequently lost the ability to bind the transcription factor (**Figure 4-2F**). This suggests ncRNAs can play a role in regulating transcription factors and, therefore, long term transcriptional regulation via SNP generation at specific target loci. While DNA methylation was observed to occur at the sh167 target site via ChIP (**Figure 4-2B**), it is not known whether there is any methylation at or surrounding the site where the SNP arose. This could be further investigated using sequencing-based methods to detect DNA methylation, which would allow for the identification of specific sites of methylation in the genome. It is also possible that the observed SNP detected at UBC22 may have already existed in a sub-population of cells, which could have been selected for during MDB4 and SMUG1 knockdown. This warrants further investigation.

MBD4 and SMUG1 knockout HAP1 cell lines were tested with both sh167 and *PTENpg1* α ncRNAs, and also demonstrated an ability to generate C to T SNPs. The sh167 shRNA was considerably more potent than the *PTENpg1* α lncRNA at inducing SNPs (~2.11-fold increase vs ~330-fold increase respectively). Notably, sh167 induced a different SNP to that observed when RNAi was used on MBD4 and SMUG1. All of the induced SNPs were located ~200bp downstream from the ncRNA target sites (**Table 4-1**), which corresponds to the length of a nucleosome. This suggests nucleosome positioning may play a role in where these SNPs occur in relation to the target locus.

Annotation of SNP sites discovered from bioinformatic analyses of ΔMBD4 and ΔSMUG1 HAP1 cells revealed that the majority of endogenously generated C to T SNPs were in the 3’ UTRs of genes, and directly 5’ of repetitive microsatellite sequences. The frequency of these repetitive microsatellite sites suggests ncRNAs may play a role in generating microsatellite indels via expansions and deletions. MBD4 mutation was previously associated with microsatellite instability in colorectal cancer (Evertson et al. 2003).

The observations presented here suggest that both lncRNAs and shRNAs are capable of directing the generation of C to T mutations at target loci when repair enzymes MBD4 and SMUG1 are repressed. This provides evidence that ncRNAs can function as mediators of genomic mutation that selection can act upon. Therefore, these mechanisms may drive evolution of the genome (Morris 2015). This may prove pivotal to understanding how content of the genome has changed over time. It might also provide the understanding required to possibly use ncRNAs as a therapeutic to revert mutations, present in various disease states, to their non-diseased nucleotide composition.

4.3 Supplemental material that accompanies the manuscript draft

4.3.1 Materials and methods

4.3.1.1 Cell cultures

Human HEK293 and 293-167 (Tet inducible shRNA UBC167 expressing (Hawkins et al. 2009a)) cell lines, commercially available wildtype HAP1 and HAP1 cells with CRISPR directed mutations in SMUG (Catalog HGT5501432, ΔSMUG) and MBD4 (Catalog HGT5500578, ΔMBD4)(ThermoFisher Scientific), or 293ΔMBD4 HEK cells, generated to contain a deletion in MBD4 (City of Hope, CRISPR engineering core facility) were utilized in the studies presented here. The various cell lines were either transduced with lentiviral vectors and FACS sorted, or transfected with plasmid expressing shRNAs and control parental pU6M2 expressing plasmids (Amarzguioui, Rossi, and Kim 2005) at a ratio of ~1µg/10^6 cells as described in (Ackley et al. 2013). Transfection of 293 based cells was carried out using Lipofectamine 3000 while transfection of HAP1 cells relied on Neon Electroporation (3µg/10^6 cells, 31575V, 10ms, 3 pulses). Transfected cultures were assessed between 24-72 hours post-transfection and gene expression determined by qRT-PCR.

4.3.1.2 RNAi of deamination pathway associated genes

Several genes involved in deamination and repair processes in human cells were subjected to targeting with RNAi. Small-hairpin RNAs (shRNAs) were generated and expressed from the U6M2 vector system (Amarzguioui, Rossi, and Kim 2005; Ackley et al. 2013). These shRNAs and their targets (shMBD4-3: targeted to Methyl-CpG (mCpG) binding domain protein 4 (MBD4) (Kondo et al. 2005), shTDG3: targeted to Thymine DNA Glycosylase (TDG) (Kunz et al. 2009), and shSMUG1: targeted to SMUG1) were screened in 293-167 cells (stable sh167 expressing HEK293 cells, treated with and without

Tetracycline drug selection for the shRNA) for suppression of their targeted gene relative to cultures treated with the parental plasmid pU6M2 (Control) as the control.

4.3.1.3 Chromatin Immunoprecipitation analysis

ChIP was carried out on 293-167 cells treated +/- Tet (1µg/ml every 2-3 days) for 3 to 7 days and PG1 and Mock cell lines for MBD4 (Abcam, ab12187), SMUG (Santa Cruz, sc-271580), APOBEC3A (Santa Cruz, sc-130688), DUT (Abcam, ab137097), UNG (Abcam, ab23926) and 5-Methylcytosine (Abcam, ab10805), SP1 (Abcam, ab13370) and TDG (Proteintech Group, 13370-1-AP) following previously established protocols (Johnsson, Ackley, et al. 2013b, 2013a). Enrichment of each particular protein was determined at either the *UBC* or *PTEN* promoter with designated primers (**Table S4-3**).

4.3.1.4 Western Blot

A total of $2x10^6$ HAP1, ΔMBD4, HEK293, or stable ΔMBD4-HEK293 cells were pelleted and stored at -80°C. Frozen cell pellets were re-suspended in 50µL M-PER Mammalian Protein Extraction Reagent (Thermo Scientific, 78501) plus Halt Protease Inhibitor Cocktail (100x) (Thermo Scientific, 78429) on ice and vortexed. Protein amount was quantitated using Pierce BCA Protein Assay Kit (Thermo Scientific, 23225). A total of 25µg of protein was mixed with 4x Laemmli (BioRad, 1610747) plus β-Mercaptoethanol, heated to 100°C and placed on ice. Samples were loaded onto 7.5% SDS gel (BioRad, 4561024) in Tris/Glycine/SDS Buffer in Mini Protean gel tanks (BioRad, 1658004) with Kaleidoscope ladder (BioRad, 1610375). Run at 100V for 10 mins followed by 200V for 20 mins. Transfer bands to nitrocellulose membrane Trans-Blot Turbo Transfer Pack (BioRad, 1704158) using Trans-Blot Turbo Transfer System (BioRad, 17001918). Run for 10 mins at 25V and 1.3A. Put membrane on shaker covered in TBST plus 5% skim milk for 1hr. 1µg/mL primary antibody was added to TBST plus 5% skim milk and shaken overnight at 4°C. Wash

3 times in TBST, add appropriate secondary antibody in 1:10,000 dilution for 1hr shaking at room temp., wash a further 3 times in TBST and image on Chemidoc Touch Imaging System (Thermo Fisher) after Pierce ECL Plus Substrate (Thermo Fisher, 32132).

4.3.1.5 qRT-PCR

qRT-PCR analysis was performed following established protocols (Lister, Clemson, and Morris 2015).

4.3.1.6 Design of Barcoded Oligos for Deep Sequencing

Promoter sequence surrounding target RNA's of interest in genes *PTEN* and *UBC* were obtained from UCSC genome browser. PCR amplicons were designed up to 1,000bp upstream and downstream of all target RNA's in that genes promoter. Amplicons were made with overlapping ends and of sizes 200-420bp, while also ensuring that target RNA's did not fall at amplicon ends but rather in the middle of an amplicon.

4.3.1.7 Detection of Single Nucleotide Polymorphisms using deep sequencing

To determine the emergence of C->T and G->A mutations arising at CpG sites in particular non-coding RNA targeted loci in the *UBC* and *PTEN* promoter loci, both RNA and genomic DNA from the various differential treated cultures were collected and MiSeq analysis performed (see Treatments 1 and 2 below).

4.3.1.8 Treatment 1: LncRNA and shRNA directed SNP induction experiment

Lentiviral vectors pHIV7-IMPDH-U6-pg1, pHIV7-IMPDH-U6-sh167 and pHIV7-IMPDH-U6-Mock were generated and calcium phosphate transfected into HEK293 cells using 15µg vector, 10µg Gag/Pol, 5µg VSVG and 4µg Rev per $4x10^6$ cells. Media supernatant was collected and sterile filtered through 0.45µm filter. Lentiviral containing media was mixed with fresh media in 1:1 ratio and added onto ΔMBD4 HAP1 cells. 12 hours

later lentiviral supernatant was removed and fresh media was added to cells. These transduced cells were then GFP FACS sorted, expanded and cultures synchronised. Transient transfections with the same vectors were also performed, transfecting on days 0, 3 and 5. On day 7 for both transient transfected cells and synchronised lentiviral transduced cells, DNA and RNA isolated to confirm repression of MBD4 and SMUG and determine any SNPs at the *PTEN* promoter locus by barcoded primer PCR (**Table S4-4 and S4-5**) and MiSeq analysis (COH Integrative Genomics Core).

4.3.1.9 Treatment 2: Whole exome sequencing in ΔMBD4 and ΔSMUG cells

HAP1 parental, ΔMBD4 HAP1, ΔSMUG HAP1 and ΔMBD4 HEK293 cell lines were synchronized, cultured for 72 hours and whole genome SNP analysis carried out on genomic DNA extracted using QIAamp DNA Mini Kit (Qiagen, 51306) (whole exome sequencing, COH Integrative Genomics Core).

4.3.2 Supplemental data

Table S4-3 Primers and shRNAs used in analysis

Primer	Sequence
MBD4_F1un	CAGTAACTTACCGGAGGTCATTC
MBD4_R1un	GACTCAGACTCTCCAGCCCA
TDG_F1un	GTTGAGAGCGTGGAGTTAAGAG
TDG_R1un	AAGAACAGGAAGAAGAAAGCCA
SMUG1_F1un	AGCCAAGCATCCACCTAGAA
SMUG1_R1un	CAGTAGCGAGTCACGTAGTTG
shMBD4-3	CCACGAACACAGATAGAAAGAA
shTDG-3	AACAGTTGTCCTTGCAAATAGC
shSMUG1-1	GCAGTAGCGAGTCACGTAGTTG
UBCproF1	AGAATCGCCGAGAAGGGACTACTT
UBCproR1	TCATAAGACTCGGCCTTAGAACCC
POSTN F	GAGCTTTACAACGGGCAAATAC
POSTN R	CTCCCTTGCTTACTCCCTTTC
PTENF3	AGAAAGCTTACAGTTGGGCCCTGT
PTENR3	GCCACAGCAAAGAATGGTGATGCT
UBC_Exon1_F	CCACAAACAAGAACTGCGACCCAA
UBC_Exon1_R	AAGTAGTCCCTTCTCGGCGATTCT
RPL10 F	CCTCTTTCCCTTCGGTGTG
RPL10 R	AATCTTGGCATCAGGGACAC

Table S4-4 Barcoded Oligonucleotides used in SNP analysis

Bold indicates barcode sequence in forward (F) primer. Stable vs transient samples were submitted separately from one another.

Primer	Sequence
UBCF2_trmt_ΔMBD4_F	**ATCACG**ATCGCCGAGAAGGGACTACTTTTCCTCGCCTGTTCCGC
UBCF3_trmt_ΔMBD4_F	**ATCACG**CTTGGCGGTCTCTCCACACGCCTCCG
UBCF4_trmt_ΔMBD4_F	**ATCACG**GTCTCTCCACACGCCTCCGTCCCACCG
UBCF5_trmt_ΔMBD4_F	**ATCACG** ACCCAAGTCCCGTCCTAAAA
UBCF2_ctrl_ΔMBD4_F	**CGATGT**TCGCCGAGAAGGGACTACTTTTCCTCGCCTGTTCCGC
UBCF3_ctrl_ΔMBD4_F	**CGATGT**CCTTGGCGGTCTCTCCACACGCCTCCG
UBCF4_ctrl_ΔMBD4_F	**CGATGT**TCTCTCCACACGCCTCCGTCCCACCG
UBCF5_ctrl_ΔMBD4_F	**CGATGT**ACCCAAGTCCCGTCCTAAAA
UBCF2_trmt_ΔSMUG1_F	**TTAGGC**ATCGCCGAGAAGGGACTACTTTTCCTCGCCTGTTCCGC
UBCF3_trmt_ΔSMUG1_F	**TTAGGC**CCTTGGCGGTCTCTCCACACGCCTCCG
UBCF4_trmt_ΔSMUG1_F	**TTAGGC**GTCTCTCCACACGCCTCCGTCCCACCG
UBCF5_trmt_ΔSMUG1_F	**TTAGGC**ACCCAAGTCCCGTCCTAAAA
UBCF2_ctrl_ΔSMUG1_F	**TGACCA**ATCGCCGAGAAGGGACTACTTTTCCTCGCCTGTTCCGC
UBCF3_ctrl_ΔSMUG1_F	**TGACCA**CCTTGGCGGTCTCTCCACACGCCTCCG
UBCF4_ctrl_ΔSMUG1_F	**TGACCA**GTCTCTCCACACGCCTCCGTCCCACCG
UBCF5_ctrl_ΔSMUG1_F	**TGACCA**ACCCAAGTCCCGTCCTAAAA

UBCF2_R	CATAAGACTCGGCCTTAGAACCCCAGTATC
UBCF3_R	CTGCGGAGGGATCTCCGTGGGGCGGT
UBCF4_R	CTGCCACGTCAGACGAAGGGCGCAGCGAGCGTCC
UBCF5_R	AAATAAAGACCCGTCCATCTCG
PTEN1_trmt_ΔMBD4_F	**ATCACG**TCGCACCCAGAGCTACCGCT
PTEN2_trmt_ΔMBD4_F	**ATCACG**ATGTGGCGGGACTCTTTATGCGCTGCGGC AGGATACGCGC
PTEN5_trmt_ΔMBD4_F	**ATCACG**TCGCCTCCTCTTCGTCTTTTCTAACCGTGC
PTEN6_trmt_ΔMBD4_F	**ATCACG**CTCGGCTTCTCCTGAAAGGGAAG
PTEN1_ctrl_ΔMBD4_F	**CGATGT**TCGCACCCAGAGCTACCGCT
PTEN2_ctrl_ΔMBD4_F	**CGATGT**ATGTGGCGGGACTCTTTATGCGCTGCGGC AGGATACGCGC
PTEN5_ctrl_ΔMBD4_F	**CGATGT**TCGCCTCCTCTTCGTCTTTTCTAACCGTGC
PTEN6_ctrl_ΔMBD4_F	**CGATGT**CTCGGCTTCTCCTGAAAGGGAAG
PTEN1_trmt_ΔSMUG1_F	**TTAGGC**TCGCACCCAGAGCTACCGCT
PTEN2_trmt_ΔSMUG1_F	**TTAGGC**ATGTGGCGGGACTCTTTATGCGCTGCGGC AGGATACGCGC
PTEN5_trmt_ΔSMUG1_F	**TTAGGC**TCGCCTCCTCTTCGTCTTTTCTAACCGTGC
PTEN6_trmt_ΔSMUG1_F	**TTAGGC**CTCGGCTTCTCCTGAAAGGGAAG
PTEN1_ctrl_ΔSMUG1_F	**TGACCA**TCGCACCCAGAGCTACCGCT
PTEN2_ctrl_ΔSMUG1_F	**TGACCA**ATGTGGCGGGACTCTTTATGCGCTGCGGC AGGATACGCGC
PTEN5_ctrl_ΔSMUG1_F	**TGACCA**TCGCCTCCTCTTCGTCTTTTCTAACCGTGC
PTEN6_ctrl_ΔSMUG1_F	**TGACCA**CTCGGCTTCTCCTGAAAGGGAAG

PTEN1_R	CGGCTCAACTCTCAAACTTCC
PTEN2_R	GAGGCTGCACGGTTAGAAAAGACGAAGAGGAGGC G
PTEN5_R	GCCCTGGAAATGGTGACAGGCGACTCAG
PTEN6_R	GGCGCACGGGAGGTTTAAAAC

Table S4-5 Amplicons from oligonucleotides used in SNP analysis

Forward and Reverse Oligonucleotides shown in capitals.

Name	Length (bp)	Amplicon Sequence
UBC2	131	ATCGCCGAGAAGGGACTACTTTTCCTCGCCTGTTCCG Ctctctggaaagaaaaccagtgccctagagtcacccaagtcccgtcctaaaatgtccttc tgctGATACTGGGGTTCTAAGGCCGAGTCTTATG
UBC3	221	CCTTGGCGGTCTCTCCACACGCCTCCGtcccaccgagcggccc ggcggccacgaaagccccggccagcccagcagcccgctactcaccaagtgacgatc acagcgatccacaaacaagaactgcgacccaaatcccggctgcgacggaactagctg tgccacacccggcgcgtccttatataatcatcggcgttcACCGCCCCACGGA GATCCCTCCGCAG
UBC4	420	GTCTCTCCACACGCCTCCGTCCCACCGagcggcccggcggcc acgaaagccccggccagcccagcagcccgctactcaccaagtgacgatcacagcgat ccacaaacaagaactgcgacccaaatcccggctgcgacggaactagctgtgccacac ccggcgcgtccttatataatcatcggcgttcaccgccccacggagatccctccgcagaa tcgccgagaagggactacttttcctcgcctgttccgctctctggaaagaaaaccagtgcc ctagagtcacccaagtcccgtcctaaaatgtccttctgctgatactggggttctaaggccg agtcttatgagcagcgggccgctgtcctgagcgtccgggcggaaggatcaGGAC GCTCGCTGCGCCCTTCGTCTGACGTGGCAG
UBC5	401	ACCCAAGTCCCGTCCTAAAAtgtccttctgctgatactggggttctaag gccgagtcttatgagcagcgggccgctgtcctgagcgtccgggcggaaggatcagga cgctcgctgcgcccttcgtctgacgtggcagcgctcgccgtgaggaggggggcgccc gcgggaggcgccaaaacccggcgcggaggccgctgcaccgggagcgggtgggcg gcccgcgttccttagccgcggctccgcggcccctcccctcagcccccgcggtccctc gggtcgcgaaccccgcgttcggcggacgtggggcacgcagcgagcgccttctgcag

		agtcggggtctccctcccctccctcctgacccgccctggcccaccctgCGAGATG GACGGGTCTTTATTT
PTEN1	293	TCGCACCCAGAGCTACCGCTctgccccctcctaccgcccccctgccct gccctgccctcccctcgcccggcgcggtcccgtccgcctctcgctcgcctcccgcctcc cctcggtcttccgaggcgcccgggctcccggcgcggcggcggaggggcgggcag gccggcgggcggtgatgtggcgggactctttatgcgctgcggcaggatacgcgctcg gcgctgggacgcgactgcgctcagttctctcctctcggaagctgcagccatgatGGA AGTTTGAGAGTTGAGCCG
PTEN2	300	ATGTGGCGGGACTCTTTATGCGCTGCGGCAGGATACG CGCtcggcgctgggacgcgactgcgctcagttctctcctctcggaagctgcagccat gatggaagtttgagagttgagccgctgtgaggcgaggccgggctcaggcgagggag atgagagacggcggcggccgcggcccggagcccctctcagcgcctgtgagcagccg cgggggcagcgccctcggggagccggccggcctgcggcggcggcagcggcggcg tttctCGCCTCCTCTTCGTCTTTTCTAACCGTGCAGCCTC
PTEN5	219	TCGCCTCCTCTTCGTCTTTTCTAACCGTGCagcctcttcctcg gcttctcctgaaagggaaggtggaagccgtgggctcgggcgggagccggctgaggc gcggcggcggcggcggcacctcccgctcctggagcgggggggagaagcggcggc ggcggcggccgcggcggctgcagctccagggaggggtCTGAGTCGCCT GTCACCATTTCCAGGGC
PTEN6	287	CTCGGCTTCTCCTGAAAGGGAAGgtggaagccgtgggctcgggc gggagccggctgaggcgcggcggcggcggcggcacctcccgctcctggagcggg ggggagaagcggcggcggcggcggccgcggcggctgcagctccagggaggggt ctgagtcgcctgtcaccatttccagggctgggaacgccggagagttggtctctccccttc tactgcctccaacacggcggcggcggcggctggcacatccagggacccgggccgG TTTTAAACCTCCCGTGCGCC

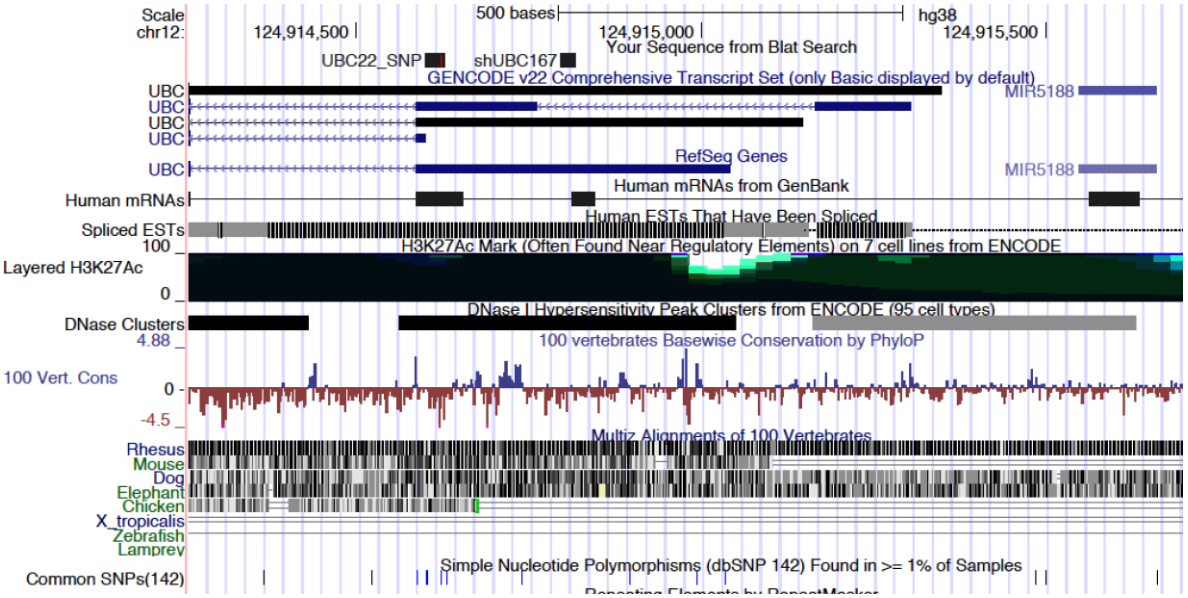


Figure S4-4 SNP found in *UBC* promoter following shMBD4-3 and shSMUG1-1 treatment in 293-167 cells

Snapshot from UCSC genome browser. SNP is found to occur ~200bp downstream of the sh167 target site in the *UBC* promoter following tetracycline treatment. Common SNPs track on UCSC genome browser showed a positional match for the SNP generated.

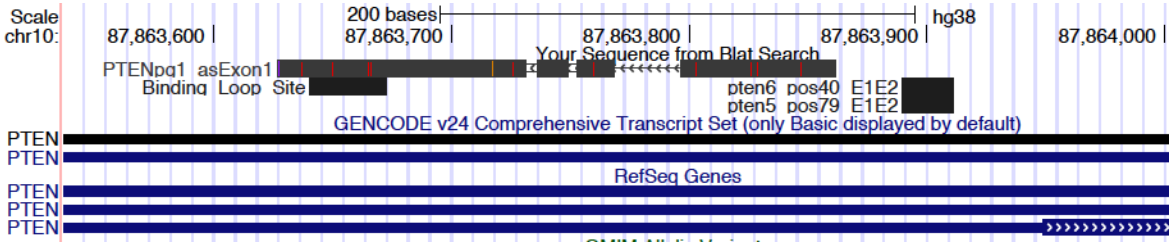


Figure S4-5 SNP found in *PTEN* promoter following *PTENpg1* α treatment in either ΔMBD4 or ΔSMUG HAP1 cells

Snapshot from UCSC genome browser. SNP is found to occur ~200bp downstream of the *PTENpg1* α binding loop site in the *PTEN* promoter. Red lines indicate nucleotide differences between *PTENpg1* asExon 1 and *PTEN* promoter sequences.

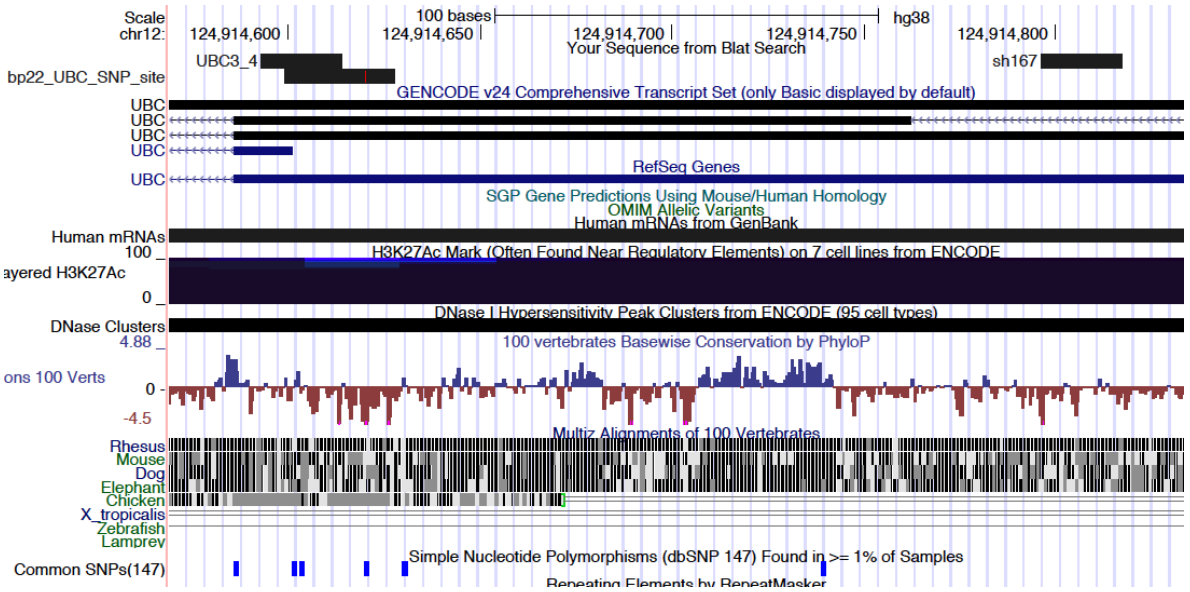


Figure S4-6 SNP found in *UBC* promoter following sh167 treatment in either ΔMBD4 or ΔSMUG HAP1 cells

Snapshot from UCSC genome browser. SNP is found to occur ~200bp downstream of sh167 target site in the *UBC* promoter. Common SNPs track on UCSC genome browser showed a positional match for the SNP generated.

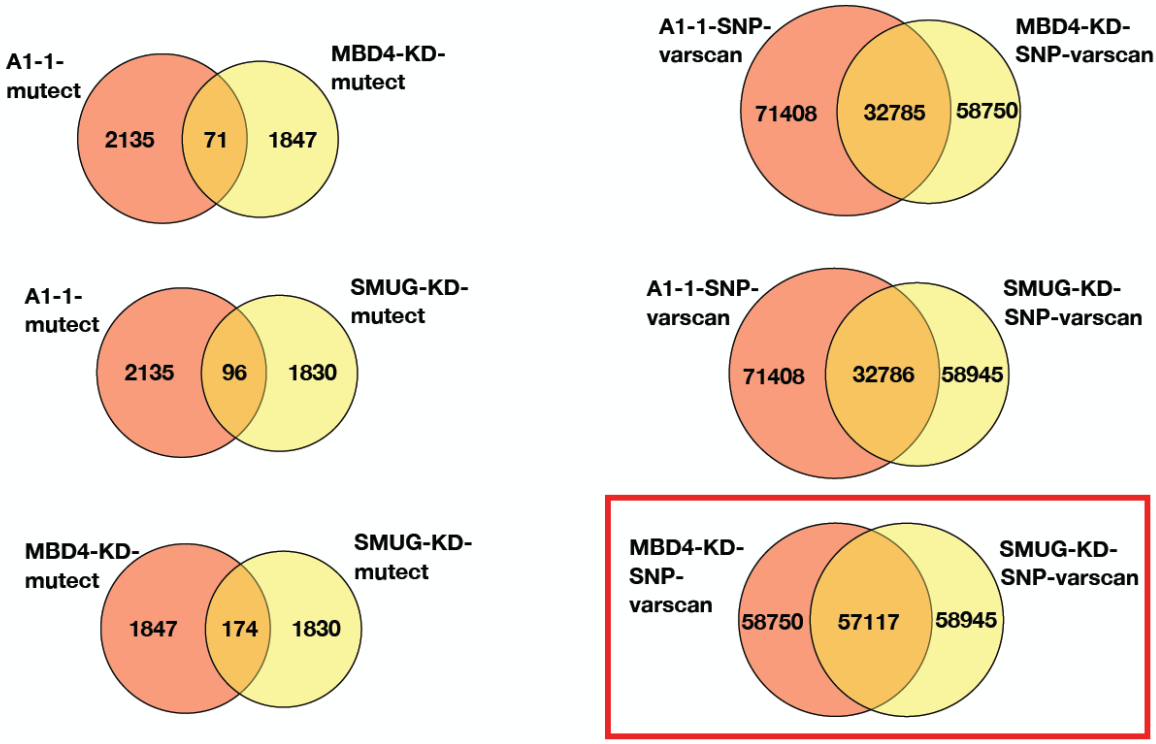


Figure S4-7 Varscan and MuTect 2 data analysis on SNPs found in ΔMBD4 and ΔSMUG cell lines

Bioinformatics data overlay of SNPs detected by either Varscan or MuTect 2 analysis from Treatment 2. Number of common SNPs between cell lines found in intersect of two circles, with number of unique SNPs to individual cell lines found in individual circles.

Motif flanking SNP, underlined	SNP location highlighted in red
GAGCAAATGTAGGTTGCA*GATTCAGGC* ***AA AAC ACACACACACACACACACAAAA*** *CCCTGAAGAGCACAATG* AGCATTATT	
CAGGTGTTCATGA**TTTCTTTTATTCTTTG**CACA**ACAACAACAACAACAAA**TTCACGGAAT	
TGTACCCATTCT**CATTCATTTTCTTCATTCCCT**TAC**ACACACACACACACC**TCTCCCTTCCGCATA	
ACTTTCTGCAATGGGATAAACTCCACTTCCTATACA**CACACACACACACACACA**TCCCCCTTGTTTGGGT	
GCTGTAATCATTAATTGTTGGCTTCAAATGTGGACA**CACACACACACACACA**TCATGCCAAGGAGGGAAT	
GGTTTTCTTTAAAAACATCAGAGCTGAATTCCTTCTAAAAT**ACA**ACA**ACAACAACAACAACAACAAAA**TAAGTACACTTGGTACCTT	
ACTTTCATTTGTTTCATGTTACCGTTTACAACTTTAATG**CACA**CAC**ACACACACACAAAA**GGATATCTACTGCAATAGAAATA	
ACGTACATATGGACCCATACACATAAGTG**CAC**ACACACACACACACACA**GAATGACAGAAAG**	
TTCCATGACAGTAGCATACTGATGAG**ACA**ACA**CACACACACACACAAAACAACA**GCAACAAC	
TATTTGTTTATGTTATTTAAAT**ACACACCCAGAC**ACA**CACACACACACA**GTTTTTGCTTTTTCTGGAAA	
GCCCACATCTGTAC**AAAACAACA**TAAGCAACA**CACACACACACACA**TATATATATGCATATA	
AATATAAGCATATCAC**AAAAACTAAAGA**ATA**AATATATATATATATATAAACATA**GGCCCTGCTGAG	
TTAAGAGAGGTAATCTCTCTATCCCTTT**GTGTGTGTGTGTGTATATATATATATAT**CATACATAATCCCATATCTATGCATCT	
AATAGCAGTGAATTTTAGTCATTAACTTC**AAA**ATA**TATATATATATATAATTTAAAGGAAAAA**TAG	
TAGTCCAGTCACAAATCTTTGTGTTGTTAATGGATTATTCAAGATA**TATATATATATTCTTA**CTACTGCATAGTTCCTCTCATAAGTCC	
CACTGCACTGCAGCCTGGGCAACATAGCAAGACTCTGTCTCA**A**AAT**AATAATAATAATAATAATAAA**GGCGTTGTTAGCTTGTAAGGAG	
TGTGA**TTTTATTTTT**AAGCTACTCCCTCTCATTC**AT****ATATATATATATATAT**GTCTATCTCCATAATTAG	
TTAACACACCCAGAAAGGCAGCTAACAGC**TAT****ATATATATATATAAATTT**CAGCCCAAACTC	
GAAATATC**TTTGTCCATAGTGGTGGTGGA**GTC**TCTCTCTCTCTCTCTCTTTTT**GTTTGCT	
AAAGCAC**TATATCTATCCTTCCATATACTGTATAT**ATT**ATAGTTCTATCTTGTT**ACACCAAATCTCACATGCAGAGAGACT	
TTTGATTCATTCGAGACAAC**ATGTGTGTA**TGT**GTGTGTGTGTGTGTGTGTA**GATAATGTGGT	
ATCCAAA**TTTTTTCTTTTGTT**ATAAAAGTCCATTACATGAGGCGTG**TGTGTGTGTGTGT**GTGTTGCTTAAAATATAATTAACACACACATC	
ATAAATCCA**TTTGATTT**GAATAGTGTGTGTGTGTACATG**GTG****TGTGTGTGTGTGTGTG**CACTTTAGTTTCGTGAGCCTTGGGA	
AAACTTGACTTCCCATGATGGAGGTTGTAAGA**GTG****TGTGTGTGTGT**AGGTATGTTTGCATGTGTTTG	
ATGCACTATTATGTATTGATTCTCCAT**GAGAC**AGA**GAGAGAGAGAGAGA**CTATCAGATAGTTTACACCC	
GATTGAACCAGAAAAT**GAGAGGGGGGGA**GGA**GAGAGAGAGAGAGAG**CTAGAAATTGAAAC	
CAATACTGTTTCTAATCTCCTCCACCT**AG**GAG**AGAGAGAGAG**TCATTACAGCATGGCAAA	
ACAAC**GGGG**TGTCA**GGGG**AGTGA**GGGG**TGGAGGTA**GGGGCCA**GAG**GGGC**TTACACGATATCCCATACCTTTAATGCCTTTGGCCTTCCAT	
AACCAGAGGGAGAGAGAGAGACAGCGAT**AG**AGA**GAGAGAGAGAGAGA**CAGAAAAGGGGGG	
AGCCATTATACCCAG**AATAAAATAATACATTTATGCTGGATTTT**ATTCAGACC**AAACTAAAAT**GGATTTGTGATGATTTGTGATTTGGTA	
GATTACCACAGTTAAATTAATCAATGTAACTATTAATGTACTAGTTACTC**TGTGTGTGT**ATATGGGAG	

Figure S4-8 Candidate SNP sites and 3' microsatellite sequences

Sequences surrounding manually annotated SNP sites found using IGV and UCSC genome browser. SNP sequences are shown in red, with microsatellite repetitive sequences in bold found 3' of the SNP.

Figure S4-9 Sense and antisense transcripts found at candidate SNP site

Snapshot from UCSC genome browser. Sense and antisense transcripts exist around SNP found at SOCS5 gene promoter. SOCS5 is a member of the Suppressor of Cytokine Signalling (SOCS) family, though its specific protein function is not known.

5 Discussion

Chapter contents:

Page No.

List of figures:

Page No.

5.1 Thesis overview

In the two decades since the completion of the human genome project, our understanding of the essential and complex roles ncRNAs play in regulating the genome has exploded. However, we have still only scratched the surface of the ncRNA world. With over ten thousand identified ncRNAs to date, only a handful have been mechanistically studied. Non-coding RNAs are integral to regulation of the genome and disease states. They are involved in a variety of functions in the cell, such as cancer progression, epigenetic regulation, chromosomal organisation, X-chromosome inactivation and regulation of the cell cycle. Here, I discuss my findings and their contribution to improve our understanding of the field of ncRNA biology, in particular the ability of ncRNAs to epigenetically modify the genome and regulate gene expression.

I examined three different types of ncRNAs over the course of this thesis. These are as6 (which is a sasRNA) targeted to the *Periostin* promoter (**Chapter 2**), *PTENpg1* α (a lncRNA) targeted to the *PTEN* promoter (**Chapters 3-4**), and sh167 (a shRNA) targeted to the *UBC* promoter (**Chapter 4**). I performed functional assays to determine the mechanisms employed by these ncRNAs to modulate or alter the genome at their target loci. In this chapter, I will develop a unifying model for the epigenetic modification of the genome and regulation of gene expression by ncRNAs that encompasses the three models described in **Chapters 2-4**. This all-encompassing model will provide insight into the complex roles ncRNAs perform in the cell. Finally, I will discuss future directions for the work (presented in this thesis) necessary to fill critical gaps in current scientific knowledge on ncRNA biology.

5.2 SasRNA directed epigenetic regulation of the genome

5.2.1 Designing sasRNAs for TGS

Antisense sncRNAs can be designed (with computational algorithms) to epigenetically modulate transcriptional genes silencing (TGS). These RNAs can direct epigenetic machinery to target loci to reduce mRNA expression, and represent a new method for therapeutically correcting improperly expressed genes (Ackley et al. 2013). This method permits stable, long-term suppression of gene expression that is somatically heritable in daughter cells (Morris 2009c; Yu et al. 2008). In this thesis, I tested the ability of a specific sasRNAs (as6) to direct TGS to the gene promoter of *Periostin*, which is involved in increased tumour metastatic and invasive potential (**Chapter 2**).

5.2.2 Periostin expression in different cancer types

Periostin overexpression was reported in a number of cancer types, including breast, ovarian, lung, prostate cancers and neuroblastoma (Kudo et al. 2007; Tischler et al. 2010; Ruan, Bao, and Ouyang 2009). When examining the *in vitro* expression (via qRT-PCR) of Periostin, only PC3 (prostate) and HEK293 (human embryonic kidney) cells demonstrated any expression of the gene. In contrast, HeLa (cervical) and CFPAC cells (pancreatic adenocarcinoma cells from cystic fibrosis patient) did not express *Periostin* (**Figure 2-1b**). As *Periostin* plays a role in increasing proliferative ability, as well as metastatic potential, this raises questions as to why HeLa and CFPAC cell lines had no observable expression of *Periostin*. Future experiments looking at the effectiveness of as6 in other cell lines that do express *Periostin* are necessary to determine if this novel TGS inducing ncRNA has more widespread therapeutic appeal.

5.2.3 CSCs and PTGS

High numbers of cancer stem cells (CSCs), which derive from cancer and have been reported in many human tumour types, correlate with increased metastatic and invasive potential (Malanchi and Huelsken 2009). Both CSCs and overexpression of *Periostin* cause long term tumour survival (Malanchi et al. 2012). It remains unknown if *Periostin* plays a role in the formation of these cells, or if reduced *Periostin* expression affects CSCs. Phenotypic data provided evidence that knockdown of *Periostin* was able to repress the cells ability for migration and proliferation (**Figure 2-3**). However, *in vitro* studies of CSCs remain challenging, as culturing CSCs causes them to lose their stem cell properties over time (Huang and Rofstad 2017), so the effect of *Periostin* expression (or suppression) remains obscure.

The data in this thesis suggests that TGS is occurring at the *Periostin* promoter, but it appears that PTGS (post transcriptional gene silencing) is also occurring, as observed in the nuclear run-on and inhibition of DNMT3a, HDAC and RNAPII experiments (**Figure 2-2**). However, this needs further investigation to uncover the precise mechanism of PTGS. The combination of these gene silencing methods (TGS and PTGS) would produce a more effective repression of *Periostin*, with sasRNA directed TGS ensuring that protein repression remains stable and heritable in daughter cells.

5.2.4 as6 directs TGS in human cells

In summary, as6 is a novel target RNA molecular capable of inducing TGS and possibly PTGS at the *Periostin* promoter. TGS of *Periostin* results in an observable phenotype, causing reduced cellular proliferative and metastatic ability. Effectiveness of as6 for future tumour treatments would depend on whether *Periostin* expression presents in an individuals cancer, as well as the development of improved delivery systems. With lentiviral

siRNA induced TGS already reported to work *in vivo* (Suzuki et al. 2013), steps towards improving delivery systems for small RNAs will lead to possible therapeutic outcomes.

5.3 LncRNA directed epigenetic regulation of the genome

5.3.1 PTEN and PTENpg1 deletion

The findings presented in **Chapter 3** provide mechanistic insight into the lncRNA regulatory network occurring at the *PTEN* promoter, but several questions are also raised. *PTENpg1* α clearly plays an important role in regulating *PTEN* expression by forming an RNA:RNA complex with a 5' UTR promoter-associated transcript. This complex induces epigenetic changes that lead to transcriptional suppression of *PTEN,* and increased cellular proliferation.

The PTEN pseudogene has been deleted and is therefore non-existent in some forms of cancer, making it unable to regulate *PTEN* expression (Poliseno et al. 2010; Poliseno et al. 2011). PTENpg1 deletion should (in theory) lead to increased *PTEN* expression based on the model presented in **Chapter 3**, resulting in apoptosis and disruption of the cell cycle (Salmena, Carracedo, and Pandolfi 2008). Loss of both *PTENpg1* and *PTEN* has been observed in melanoma, leading to unregulated tumour progression (Poliseno et al. 2011). Further investigation is required to elucidate how, after loss of *PTENpg1, PTEN* is still epigenetically suppressed to cause tumour progression.

LncRNAs are known to be able to bind to hundreds of loci, bringing with them chromatin remodelling proteins (Chu et al. 2011). This suggests PTENpg1 asRNAs may have many other targets, and do not exclusively regulate *PTEN* expression. The ability of PTENpg1 asRNAs to regulate gene expression or act as a miRNA sponge (Poliseno et al. 2010; Poliseno and Pandolfi 2015) may also depend on miRNA or chromatin remodelling protein expression levels. Low levels of either miRNA or chromatin modifying proteins could lead to a loss of function as either a miRNA sponge or gene suppressor of *PTEN*.

5.3.2 lncRNA secondary structure

M-FOLD analysis was used to predict the secondary structure of PTENpg1 α and found the presence of several loops throughout (**Figure S3-7**). The F4R1 variant (smallest functional truncation of PTENpg1 α) only contains one of these loops, yet is observed to have the same silencing effect as full length PTENpg1 α i.e. capable of directing epigenetic regulatory proteins to the *PTEN* promoter. The F5R2 variant also contains a loop structure, however, unlike the F4R1 variant, F5R2 cannot bind both DNMT3a and the *PTEN* promoter at the same time. Loop structure and sequence appear to be integral to this process, as only a 2bp difference exist between the F4R1 and F5R2 variants. It is not apparent why one loop structure (F4R1) can bind to both protein (DNMT3a) and promoter simultaneously, yet the other (F5R2) can only bind one of these (DNMT3a or promoter) at any given time.

5.3.3 PTENpg1 epigenetic regulation of PTEN through RNA:RNA interactions

In **Chapter 3** I demonstrated that an RNA:RNA interaction is involved in *PTENpg1* asRNA exon 1 targeting of the *PTEN* promoter. The conserved *PTENpg1* domain required for this interaction consisted of a major loop domain that interacts with the nascent 5'-UTR of *PTEN*. The *PTENpg1* major loop domain appears to be required for both interacting with the *PTEN* promoter, and binding DNMT3a protein. Finally, RNA:RNA interactions, particularly at gene promoters, may have broader relevance in lncRNA epigenetic regulation of expression for other protein-coding genes, and could prove exciting in the development of targeted therapeutics.

5.4 ncRNA directed modification of the genome

5.4.1 Key deamination repair proteins

The recruitment of proteins involved in the deamination repair pathway has been reported following CpG methylation. Deamination of 5mC results in it being converted to uracil by APOBEC3A, followed by subsequent base excision removal (BER) to restore the genome to its original nucleotide content (Wijesinghe and Bhagwat 2012; Carpenter et al. 2012; Hashimoto, Hong, et al. 2012; Jacobs and Schar 2012). Sh167 targeting of the *UBC* promoter results in an increase of CpG methylation (Hawkins et al. 2009a). As described in **Chapter 4**, there was an enrichment of key proteins involved in the 5mC deamination repair pathway at the *UBC* promoter. These include DUT, APOBEC3A, MBD4, SMUG1 and UNG (**Figures 4-2A-C**).

The protein TDG, however, was observed to have a two-fold decrease in enrichment at the *UBC* promoter (**Figure 4-2B**). Also, treatment of cultured cells with sh167, coupled with shRNA knockdown of TDG (shTDG-3, see **Table S4-3**), did not result in an observed C to T mutation like it did for shRNAs targeted to MBD4 (shMBD4-3) and SMUG1 (shSMUG1-1) (**Figure 4-2E**). It may be possible that sh167 itself interferes with the ability of TDG to perform its role in repairing the genome content. If its role in DNA repair is already hindered, in theory both control and treatment samples should develop the same SNP, and no difference would be observed between the two. Also, the discrepancy seen between the 9.3% of cells in which sh167 induced a SNP (**Figure 4-2E**) and the 87% loss of SP1 target site binding may also be due to SP1 binding interference by sh167. Further investigation is required to understand why TDG, involved in BER, is not enriched at the *UBC* promoter following sh167 treatment, as well as the large loss of SP1 binding at the SNP site. TDG is expected to have a functionally significant role following CpG methylation.

5.4.2 Nucleosome positioning, microsatellite sites and shRNAs vs lncRNAs

MBD4 was previously been reported to have a role in the generations of genomic mutations, and is thought to contribute to tumour progression when silenced (Tricarico et al. 2015). Upon examining the ability of sh167 (an shRNA) and *PTENpg1* α (a lncRNA) to generate SNPs at their target promoters (*UBC* and *PTEN* respectively) in the absence of MBD4 or SMUG1, C to T mutation was observed. Sh167 was much more potent than *PTENpg1* α (**Table 4-1**) at generating these mutations, indicating that shRNAs are more efficient at generating genomic change. The reason for this observed difference in potency between ncRNAs is unknown, and requires further investigation, as both short and long ncRNAs are capable of directing CpG methylation. The possibility also remains that shRNAs are only more potent in the cell lines that were tested in this study.

The generation of the mutations 200bp downstream of the respective ncRNA target sites indicates nucleosome position may also play an important mechanistic role at 5mC sites (**Figures S4-4-6**). Alterations to the nucleosome in future experiments may shed light on a mechanism utilised by the cell to produce C to T conversions. Many of the SNP were located in 3' UTRs throughout the genome, and always directly 5' of repetitive microsatellite sites. MBD4 was previously found to have a role in microsatellite instability (Evertson et al. 2003), suggesting that these ncRNA directed SNPs correlate with the generation of microsatellite indels. However, this requires further detailed experimentation on the generation of microsatellite indels in relation to ncRNAs before any solid conclusions could be drawn.

5.4.3 ncRNA directed genetic variation

It appears that ncRNAs are capable of editing the content of the genome by directing SNP generation at target loci. This could provide selective advantages in the event of changing evolutionary pressures. Mutations alter transcription factor binding sites following CpG deamination (Zemojtel et al. 2011) that could result in heritable changes in gene

expression. The sh167 shRNA produced a SNP at an SP1 transcription factor binding site (**Figure 4-2F**), perturbing its function. This suggests that ncRNAs (which frequently targeted promoter) direct C to T conversions that are involved in transcription factor binding site evolution (Johnsson, Ackley, et al. 2013b; Johnsson et al. 2014). However, a consequence of this might also be the loss of specific ncRNA mediated control. Change to target genomic sequence composition will result in a loss of sequence complementarity, and subsequent ncRNA targeting capabilities.

5.4.4 shRNA and lncRNA generation of SNPs

In **Chapter 4**, I demonstrated that SNPs can indeed be generated by either an shRNA (sh167) or a lncRNA (*PTENpg1* α) at their target promoters (*UBC* or *PTEN*) in the absence of MBD4 or SMUG1 proteins. This suggests ncRNAs are capable of altering the content of the genome following epigenetic rearrangement and subsequent CpG methylation of a target locus when repair pathways are hindered. Ultimately, this may provide important insight into how the cell can adapt to selective pressures on transcription factor binding sites, and may prove therapeutically relevant in the treatment of diseases in the future.

5.5 Unifying model

Here I present a model (**Figure 5-1** below) that represents all the types of epigenetic regulation and modification of the genome performed by ncRNAs described throughout this thesis. Expression of a sasRNA in the cell leads to TGS (as seen in **Chapter 2**), and expression of a *cis* or *trans* acting lncRNA within a cell leads to epigenetic modulation of a target locus (as seen in **Chapter 3**). The ncRNAs recruit epigenetic remodelling proteins, including DNMT3a, HDAC-1 and EZH2, to the target locus, leading to DNA methylation and chromatin change (Weinberg et al. 2006; Hawkins et al. 2009a; Johnsson, Ackley, et al. 2013b). This results in a transcriptionally inactive site inaccessible to RNAPII, resulting in silencing of gene expression (Morris et al. 2004; Morris and Mattick 2014a).

NcRNA modification of the genome can occur when the DNA deamination repair pathway is perturbed (as seen in **Chapter 4**). Methylated cytosines at CpG residues undergo C to U conversion by APOBEC3A (Wijesinghe and Bhagwat 2012; Carpenter et al. 2012; Suspene et al. 2013). TDG recognises and actively processes/excises lesions from deamination of C, which initiates the BER (base excision repair) pathway (Sjolund, Senejani, and Sweasy 2013; Hashimoto, Hong, et al. 2012). Other DNA glycosylases involved include MBD4 and SMUG1 (Hashimoto, Liu, et al. 2012; Kondo et al. 2005; Kemmerich et al. 2012). AP endonucleases, DNA polymerase and DNA ligase act to complete the BER process (Jacobs and Schar 2012). However, in the event that base excision repair is disrupted or blocked, the C to T conversion (uracil is recognised as a thymine residue following DNA replication) remains in the genome (Morris 2015). The generation of this SNP results from ncRNA directed modification of the genome, and can be heritably passed on to daughter cells.

Neofunctionalisation (new function for a gene following mutation) is one possible outcome in the event of germ line ncRNA directed SNP generation (in viable offspring). However, in somatic tissues, mutagenesis is more likely to result in either cellular apoptosis or tumorigenesis, depending on the location of the SNP in the genome and cellular responses.

Suppression of BER by RNAs (e.g. by RNAi) is a potential method for disruption of DNA repair, which would mean there are several layers of RNA regulation occurring simultaneously that result in the generation of a SNP. This can, in turn, disrupt the binding of the initial ncRNA that directed the SNP in the first place, which could disrupt the initial regulatory network and prevent further regulation of the target loci.

This model for epigenetic regulation of gene expression and modification of the genome is consistent with the action of other subclasses of ncRNAs. MiRNAs degrade or prevent translation of mRNAs (Fang et al. 2013; Guo et al. 2010), and siRNAs are involved in the RNAi pathway that causes PTGS (Waterhouse, Graham, and Wang 1998; Lee et al. 2003; Bernstein et al. 2001). However, antisense transcript of varying lengths (as6 is 25 bp, *PTENpg1* α is 242 bp) are able to perform transcriptional suppression through epigenetic regulatory mechanisms as presented in this model (Yu et al. 2008; Morris et al. 2008; Hawkins and Morris 2010a). Some of the sncRNA subclasses, and all of the lncRNA subclasses, can be transcribed as asRNAs (dependent upon the transcriptional activity of a given loci), and as such fit into this model. Antisense transcription is pervasive throughout the genome (Katayama et al. 2005), suggesting thousands of ncRNAs take part in this complex transcriptional system of epigenetically regulating gene expression and modifying content of the genome.

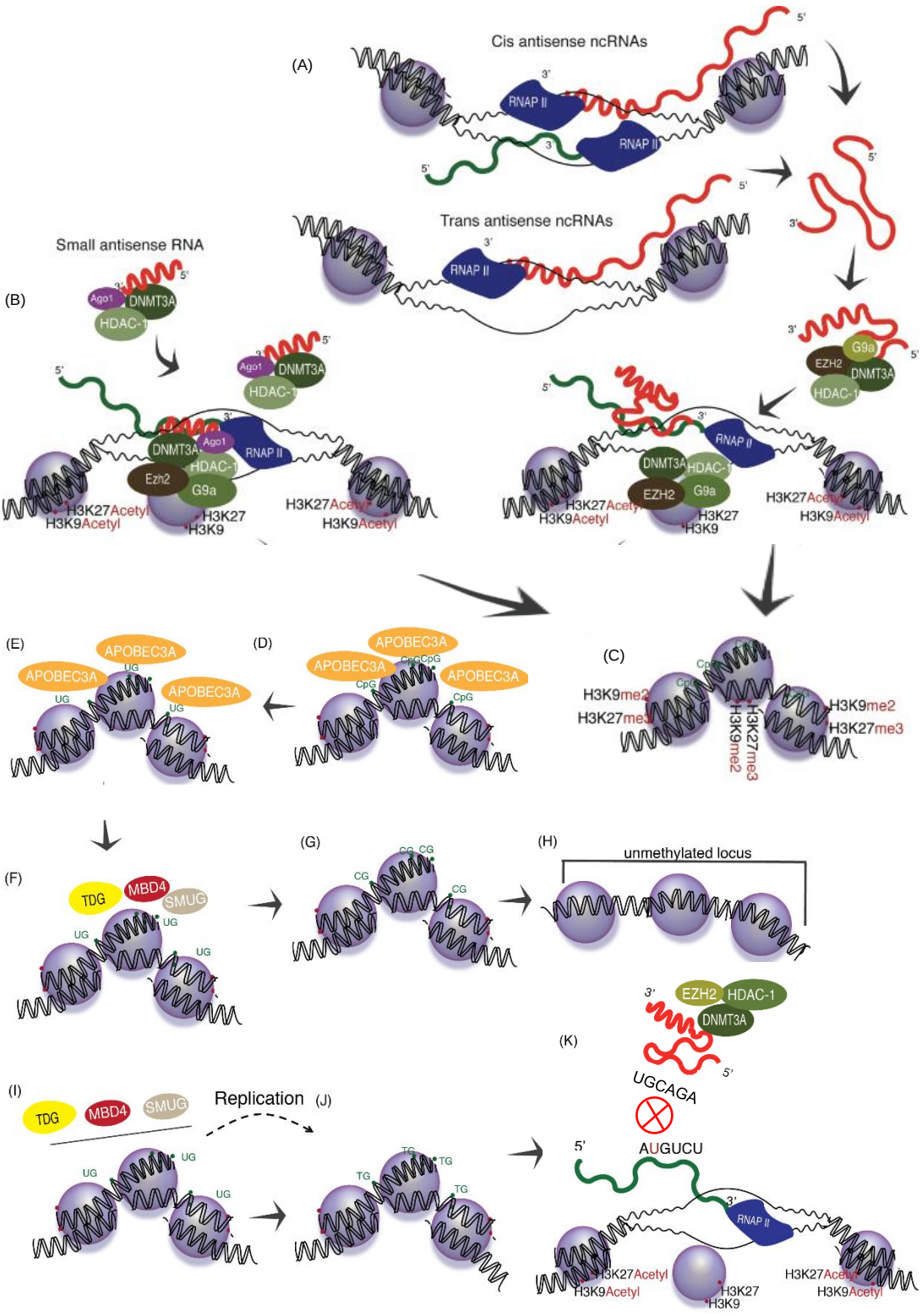

(A)
Cis antisense ncRNAs
RNAP II
Trans antisense ncRNAs
(B)
Small antisense RNA
Ago1
DNMT3A
HDAC-1
Ezh2
G9a
EZH2
H3K27Acetyl
H3K9Acetyl
H3K27
H3K9
(C)
H3K9me2
H3K27me3
(D)
APOBEC3A
CpG
(E)
UG
(F)
TDG
MBD4
SMUG
(G)
CG
(H)
unmethylated locus
(I)
Replication
(J)
TG
(K)
UGCAGA
AUGUCU

Figure 5-1 Short and long ncRNA epigenetic regulation and modification of the genome

Unifying model that depicts sncRNAs and lncRNAs directing epigenetic modification to modulate gene expression. Non-coding RNAs can also mediate the generation of SNP mutations that ultimately change nucleotide content of the genome. This becomes transgenerational if mutations occur in the germ line. (**A**) Endogenous *cis* and *trans* acting antisense lncRNA transcriptional modulation via recruitment of epigenetic silencing proteins such as DNMT3a, EZH2 and HDAC-1. (**B**) Exogenous sasRNA directed TGS that is capable of bypassing the endogenous lncRNA regulatory process. (**C**) Both endogenous and exogenous ncRNAs result in epigenetic silencing of the target locus. Epigenetic changes include histone methylation of H3K9me2 and H3K27me3, as well as CpG DNA methylation. (**D**) The methylated cytosines at (or adjacent to) the targeted RNA undergo deamination by the action of APOBEC3A, whereby (**E**) the methylated cytosine is converted to uracil (UG mismatches shown in green). (**F-H**) The resulting uracil/guanine pairing mistake is repaired by TDG, MBD4 and SMUG1. However, if this process is disrupted (**I**) (indicated by the line underneath TDG, MBD4 and SMUG1, which represents knockout or disruption of protein function) and/or replication occurs before repair of the locus is completed, then the thymine can become fixed in one daughter strand. In this event (**J**), a permanent cytosine to thymine change has occurred in the genome (TG mismatches shown in green). This could alter downstream RNA/RNA interactions (**K**) or DNA/protein or RNA/protein interactions. Adapted from (Lister, Clemson, and Morris 2015).

5.6 Future directions

The aims of this thesis were to investigate the roles ncRNAs play in the epigenetic regulation of gene expression and in the emergence of SNPs in the genome, in an effort to improve our biological understanding of RNA and its many functions inside the cell. In **Chapter 2**, TGS induced by as6 was shown to be an effective inhibitor of metastatic and invasive potential of tumour cells. However, an effective method for introduction of the sasRNA into target cells of human patients is required before it can be used to treat cancers. New delivery mechanisms that specifically target tumour cells, and do not produce unwanted side effects, are required before it is viable to treat tumours via sasRNA directed TGS. The resistance of sasRNAs to nucleases is also required, otherwise they will be degraded before inducing TGS. However, if these issues are resolved in the near future, a novel therapeutic for *Periostin* repression (and therefore metastatic inhibition) could be suggested by using as6.

The sasRNA as6 has the ability to epigenetically modify its target gene promoter to endure stable silencing (due to the heritable nature of TGS). Future experiments (via mass spectrometry) that identify the proteins associated with as6 directed TGS would prove extremely useful in determining exact TGS mechanism and discovering novel protein functions. ELISA could be used to examine Periostin protein expression outside the cell. Studying the effect of as6 in *in vitro* mice models would be valuable for gaining insight into the potential for as6 as an anti-cancer treatment (in conjunction with chemotherapeutic agents).

PTENpg1 plays an important role in the regulation of *PTEN* expression, as seen in **Chapter 3**. However, its role in the potential regulation of other loci has not been investigated. A more thorough understanding of *PTENpg1* asRNAs (α and β, emanating from the *PTENpg1* bidirectional promoter) and their binding sites throughout the genome will provide insight into the roles it plays outside of *PTEN* regulation. As *PTENpg1* β is involved

in miRNA sponging, an investigation into known miRNA binding partners may help to elucidate some of these potential binding sites.

PTEN plays an important role in tumour suppression and is frequently down regulated in cancer, so instances of *PTENpg1* and/or *PTEN* deletion in tumours should be further examined. The mechanisms for tumour formation are of interest in *PTEN* and or *PTENpg1* deletion, as *PTENpg1* deletion should lead to *PTEN* expression and subsequent apoptosis of tumour cells, whereas deletion of both genes should negate the need for epigenetic regulation of *PTEN* altogether.

In **Chapter 4**, an sasRNA and lncRNA were shown to be capable of directing C to T conversions (SNPs) in the human genome in the absence of DNA repair proteins (MBD4 and SMUG1). The locations of 43 SNPs generated in ΔMBD4 and ΔSMUG1 HAP1 cells were annotated, with most SNPs found in 3' UTRs that exhibited both sense and antisense transcription. Other members of the laboratory I carried out my PhD in are currently working to generate vectors containing some of these sense and antisense ncRNA transcripts found at the annotated SNP sites. These vectors will be transfected into either ΔMBD4 and ΔSMUG1 HAP1 cells, or HEK293 cells with MBB4 and SMUG1 suppressed by RNAi, and the target sites re-examined to determine if the same SNPs are observed.

So far, only one shRNA and one lncRNA have been tested for their ability to generate SNPs in the absence of MBD4 and SMUG1. There are a number of other subclasses of short and long ncRNAs that may be able to perform this function, which are yet to be tested. I have also generated shRNA vectors targeted to genes for *Akt*, *Myc* and *Periostin* (as the shRNA sh167 was observed to be more effective than the *PTENpg1* α lncRNA), and designed barcoded oligonucleotides across each respective promoter. These will also be transfected into ΔMBD4 and ΔSMUG1 HAP1 cells to look for SNP generation. Future experiments could also examine knockout of various other proteins involved in the deamination repair

pathway. This would determine if the absence of other DNA repair proteins results in ncRNA directed generation of SNPs. These experiments will improve our understanding of how ncRNAs play a role in editing the content of the genome, and whether this can be therapeutically utilised in the future in the treatment of diseases.

5.7 Concluding remarks

NcRNA biology has progressed at a rapid rate over the past two decades, as their vital functional roles in the cell have become more widely recognised. More recent sequencing technologies have revealed that at least 60-70% of the genome is transcribed. As a result, tens of thousands of ncRNAs have been identified, with the vast majority yet to be functionally characterized. There are likely decades of work remaining to uncover the finer detail of each of their roles in epigenetically modifying the genome and regulating gene expression. Indeed, there are perhaps roles for ncRNAs that are currently unknown. Uncovering this complex RNA world was the motivation behind this thesis, so as to improve our biological understanding of the cell and all of its processes.

6 References

Ackley, A., A. Lenox, K. Stapleton, S. Knowling, T. Lu, K. S. Sabir, P. K. Vogt, and K. V. Morris. 2013. 'An Algorithm for Generating Small RNAs Capable of Epigenetically Modulating Transcriptional Gene Silencing and Activation in Human Cells', *Mol Ther Nucleic Acids*, 2: e104.

Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, and M. R. Stratton. 2013. 'Signatures of mutational processes in human cancer', *Nature*, 500: 415-21.

Ali, I. U., L. M. Schriml, and M. Dean. 1999. 'Mutational spectra of PTEN/MMAC1 gene: a tumor suppressor with lipid phosphatase activity', *J Natl Cancer Inst*, 91: 1922-32.

Alimonti, A., A. Carracedo, J. G. Clohessy, L. C. Trotman, C. Nardella, A. Egia, L. Salmena, K. Sampieri, W. J. Haveman, E. Brogi, A. L. Richardson, J. Zhang, and P. P. Pandolfi. 2010. 'Subtle variations in Pten dose determine cancer susceptibility', *Nat Genet*, 42: 454-8.

Amarzguioui, M., J. J. Rossi, and D. Kim. 2005. 'Approaches for chemically synthesized siRNA and vector-mediated RNAi', *FEBS Lett*, 579: 5974-81.

Bao, S., G. Ouyang, X. Bai, Z. Huang, C. Ma, M. Liu, R. Shao, R. M. Anderson, J. N. Rich, and X. F. Wang. 2004. 'Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway', *Cancer Cell*, 5: 329-39.

Baril, P., R. Gangeswaran, P. C. Mahon, K. Caulee, H. M. Kocher, T. Harada, M. Zhu, H. Kalthoff, T. Crnogorac-Jurcevic, and N. R. Lemoine. 2007. 'Periostin promotes invasiveness and resistance of pancreatic cancer cells to hypoxia-induced cell death: role of the beta4 integrin and the PI3k pathway', *Oncogene*, 26: 2082-94.

Batey, R. T., M. B. Sagar, and J. A. Doudna. 2001a. 'Structural and energetic analysis of RNA recognition by a universally conserved protein from the signal recognition particle', *J Mol Biol*, 307: 229-46.

Batey, Robert T., M. Bidya Sagar, and Jennifer A. Doudna. 2001b. 'Structural and energetic analysis of RNA recognition by a universally conserved protein from the signal recognition particle1', *Journal of Molecular Biology*, 307: 229-46.

Beltran, M., I. Puig, C. Pena, J. M. Garcia, A. B. Alvarez, R. Pena, F. Bonilla, and A. G. de Herreros. 2008. 'A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition', *Genes Dev*, 22: 756-69.

Ben, Q. W., X. L. Jin, J. Liu, X. Cai, F. Yuan, and Y. Z. Yuan. 2011. 'Periostin, a matrix specific protein, is associated with proliferation and invasion of pancreatic cancer', *Oncol Rep*, 25: 709-16.

Ben, Q. W., Z. Zhao, S. F. Ge, J. Zhou, F. Yuan, and Y. Z. Yuan. 2009. 'Circulating levels of periostin may help identify patients with more aggressive colorectal cancer', *Int J Oncol*, 34: 821-8.

Bernstein, E., A. A. Caudy, S. M. Hammond, and G. J. Hannon. 2001. 'Role for a bidentate ribonuclease in the initiation step of RNA interference', *Nature*, 409: 363-6.

Brannan, C. I., E. C. Dees, R. S. Ingram, and S. M. Tilghman. 1990. 'The product of the H19 gene may function as an RNA', *Mol Cell Biol*, 10: 28-36.

Brockdorff, N., A. Ashworth, G. F. Kay, V. M. McCabe, D. P. Norris, P. J. Cooper, S. Swift, and S. Rastan. 1992. 'The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus', *Cell*, 71: 515-26.

Brown, C. J., A. Ballabio, J. L. Rupert, R. G. Lafreniere, M. Grompe, R. Tonlorenzi, and H. F. Willard. 1991. 'A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome', *Nature*, 349: 38-44.

Brown, C. J., B. D. Hendrich, J. L. Rupert, R. G. Lafreniere, Y. Xing, J. Lawrence, and H. F. Willard. 1992. 'The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus', *Cell*, 71: 527-42.

Budworth, H., and C. T. McMurray. 2013. 'Bidirectional transcription of trinucleotide repeats: roles for excision repair', *DNA Repair (Amst)*, 12: 672-84.

Camblong, J., N. Iglesias, C. Fickentscher, G. Dieppois, and F. Stutz. 2007. 'Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae', *Cell*, 131: 706-17.

Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, R. Kodzius, K. Shimokawa, V. B. Bajic, S. E. Brenner, S. Batalov, A. R. Forrest, M. Zavolan, M. J. Davis, L. G. Wilming, V. Aidinis, J. E. Allen, A. Ambesi-Impiombato, R. Apweiler, R. N. Aturaliya, T. L. Bailey, M. Bansal, L. Baxter, K. W. Beisel, T. Bersano, H. Bono, A. M. Chalk, K. P. Chiu, V. Choudhary, A. Christoffels, D. R. Clutterbuck, M. L. Crowe, E. Dalla, B. P. Dalrymple, B. de Bono, G. Della Gatta, D. di Bernardo, T. Down, P. Engstrom, M. Fagiolini, G. Faulkner, C. F. Fletcher, T. Fukushima, M. Furuno, S. Futaki, M. Gariboldi, P. Georgii-Hemming, T. R. Gingeras, T. Gojobori, R. E. Green, S. Gustincich, M. Harbers, Y. Hayashi, T. K. Hensch, N. Hirokawa, D. Hill, L. Huminiecki, M. Iacono, K. Ikeo, A. Iwama, T. Ishikawa, M. Jakt, A. Kanapin, M. Katoh, Y. Kawasawa, J. Kelso, H. Kitamura, H. Kitano, G. Kollias, S. P. Krishnan, A. Kruger, S. K. Kummerfeld, I. V. Kurochkin, L. F. Lareau, D. Lazarevic, L. Lipovich, J. Liu, S. Liuni, S. McWilliam, M. Madan Babu, M. Madera, L. Marchionni, H. Matsuda, S. Matsuzawa, H. Miki, F. Mignone, S. Miyake, K. Morris, S. Mottagui-Tabar, N. Mulder, N. Nakano, H. Nakauchi, P. Ng, R. Nilsson, S. Nishiguchi, S. Nishikawa, F. Nori, O. Ohara, Y. Okazaki, V. Orlando, K. C. Pang, W. J. Pavan, G. Pavesi, G. Pesole, N. Petrovsky, S. Piazza, J. Reed, J. F. Reid, B. Z. Ring, M. Ringwald, B. Rost, Y. Ruan, S. L. Salzberg, A. Sandelin, C. Schneider, C. Schonbach, K. Sekiguchi, C. A. Semple, S. Seno, L. Sessa, Y. Sheng, Y. Shibata, H. Shimada, K. Shimada, D. Silva, B. Sinclair, S. Sperling, E. Stupka, K. Sugiura, R. Sultana, Y. Takenaka, K. Taki, K. Tammoja, S. L. Tan, S. Tang, M. S. Taylor, J. Tegner, S. A. Teichmann, H. R. Ueda, E. van Nimwegen, R. Verardo, C. L. Wei, K. Yagi, H. Yamanishi, E. Zabarovsky, S. Zhu, A. Zimmer, W. Hide, C. Bult, S. M. Grimmond, R. D. Teasdale, E. T. Liu, V. Brusic, J. Quackenbush, C. Wahlestedt, J. S. Mattick, D. A. Hume, C. Kai, D. Sasaki, Y. Tomaru, S. Fukuda, M. Kanamori-Katayama, M. Suzuki, J. Aoki, T.

Arakawa, J. Iida, K. Imamura, M. Itoh, T. Kato, H. Kawaji, N. Kawagashira, T. Kawashima, M. Kojima, S. Kondo, H. Konno, K. Nakano, N. Ninomiya, T. Nishio, M. Okada, C. Plessy, K. Shibata, T. Shiraki, S. Suzuki, M. Tagami, K. Waki, A. Watahiki, Y. Okamura-Oho, H. Suzuki, J. Kawai, Y. Hayashizaki, Fantom Consortium, Riken Genome Exploration Research Group, and Group Genome Science. 2005. 'The transcriptional landscape of the mammalian genome', *Science*, 309: 1559-63.

Carpenter, M. A., M. Li, A. Rathore, L. Lackey, E. K. Law, A. M. Land, B. Leonard, S. M. Shandilya, M. F. Bohn, C. A. Schiffer, W. L. Brown, and R. S. Harris. 2012. 'Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A', *J Biol Chem*, 287: 34801-8.

Carracedo, A., and P. P. Pandolfi. 2008. 'The PTEN-PI3K pathway: of feedbacks and cross-talks', *Oncogene*, 27: 5527-41.

Chen, J., M. Sun, W. J. Kent, X. Huang, H. Xie, W. Wang, G. Zhou, R. Z. Shi, and J. D. Rowley. 2004. 'Over 20% of human transcripts might form sense-antisense pairs', *Nucleic Acids Res*, 32: 4812-20.

Chen, Z., L. C. Trotman, D. Shaffer, H. K. Lin, Z. A. Dotan, M. Niki, J. A. Koutcher, H. I. Scher, T. Ludwig, W. Gerald, C. Cordon-Cardo, and P. P. Pandolfi. 2005. 'Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis', *Nature*, 436: 725-30.

Cheng, J., P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, J. Long, D. Stern, H. Tammana, G. Helt, V. Sementchenko, A. Piccolboni, S. Bekiranov, D. K. Bailey, M. Ganesh, S. Ghosh, I. Bell, D. S. Gerhard, and T. R. Gingeras. 2005. 'Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution', *Science*, 308: 1149-54.

Choi, K. U., J. S. Yun, I. H. Lee, S. C. Heo, S. H. Shin, E. S. Jeon, Y. J. Choi, D. S. Suh, M. S. Yoon, and J. H. Kim. 2011. 'Lysophosphatidic acid-induced expression of periostin in stromal cells: Prognoistic relevance of periostin expression in epithelial ovarian cancer', *Int J Cancer*, 128: 332-42.

Chu, C., K. Qu, F. L. Zhong, S. E. Artandi, and H. Y. Chang. 2011. 'Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions', *Mol Cell*, 44: 667-78.

Cocquerelle, C., B. Mascrez, D. Hetuin, and B. Bailleul. 1993. 'Mis-splicing yields circular RNA molecules', *FASEB J*, 7: 155-60.

Contie, S., N. Voorzanger-Rousselot, J. Litvin, P. Clezardin, and P. Garnero. 2011. 'Increased expression and serum levels of the stromal cell-secreted protein periostin in breast cancer bone metastases', *Int J Cancer*, 128: 352-60.

Dahia, P. L., M. G. FitzGerald, X. Zhang, D. J. Marsh, Z. Zheng, T. Pietsch, A. von Deimling, F. G. Haluska, D. A. Haber, and C. Eng. 1998. 'A highly conserved processed PTEN pseudogene is located on chromosome band 9p21', *Oncogene*, 16: 2403-6.

Deraz, E. M., Y. Kudo, M. Yoshida, M. Obayashi, T. Tsunematsu, H. Tani, S. B. Siriwardena, M. R. Keikhaee, G. Qi, S. Iizuka, I. Ogawa, G. Campisi, L. Lo Muzio, Y. Abiko, A. Kikuchi, and T. Takata. 2011. 'MMP-10/stromelysin-2 promotes invasion of head and neck cancer', *PLoS One*, 6: e25438.

Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D. G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J. B. Brown, L. Lipovich, J. M. Gonzalez, M. Thomas, C. A. Davis, R. Shiekhattar, T. R. Gingeras, T. J. Hubbard, C. Notredame, J. Harrow, and R. Guigo. 2012. 'The GENCODE v7 catalog of human long noncoding RNAs:

analysis of their gene structure, evolution, and expression', *Genome Res*, 22: 1775-89.

Di Ruscio, A., A. K. Ebralidze, T. Benoukraf, G. Amabile, L. A. Goff, J. Terragni, M. E. Figueroa, L. L. D. Pontes, M. Alberich-Jorda, P. Zhang, M. C. Wu, F. D'Alo, A. Melnick, G. Leone, K. K. Ebralidze, S. Pradhan, J. L. Rinn, and D. G. Tenen. 2013. 'DNMT1-interacting RNAs block gene-specific DNA methylation', *Nature*, 503: 371-6.

Du, T., and P. D. Zamore. 2005. 'microPrimer: the biogenesis and function of microRNA', *Development*, 132: 4645-52.

Duner, S., J. Lopatko Lindman, D. Ansari, C. Gundewar, and R. Andersson. 2010. 'Pancreatic cancer: the role of pancreatic stellate cells in tumor progression', *Pancreatology*, 10: 673-81.

Eckstein, F. 2000. 'Phosphorothioate oligodeoxynucleotides: what is their origin and what is unique about them?', *Antisense Nucleic Acid Drug Dev*, 10: 117-21.

Engstrom, P. G., H. Suzuki, N. Ninomiya, A. Akalin, L. Sessa, G. Lavorgna, A. Brozzi, L. Luzi, S. L. Tan, L. Yang, G. Kunarso, E. L. Ng, S. Batalov, C. Wahlestedt, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, C. Wells, V. B. Bajic, V. Orlando, J. F. Reid, B. Lenhard, and L. Lipovich. 2006. 'Complex Loci in human and mouse genomes', *PLoS Genet*, 2: e47.

Erkan, M., C. Reiser-Erkan, C. W. Michalski, S. Deucker, D. Sauliunaite, S. Streit, I. Esposito, H. Friess, and J. Kleeff. 2009. 'Cancer-stellate cell interactions perpetuate the hypoxia-fibrosis cycle in pancreatic ductal adenocarcinoma', *Neoplasia*, 11: 497-508.

Evertson, S., A. Wallin, G. Arbman, S. Rutten, A. Emterling, H. Zhang, and X. F. Sun. 2003. 'Microsatellite instability and MBD4 mutation in unselected colorectal cancer', *Anticancer Res*, 23: 3569-74.

Faghihi, M. A., M. Zhang, J. Huang, F. Modarresi, M. P. Van der Brug, M. A. Nalls, M. R. Cookson, G. St-Laurent, 3rd, and C. Wahlestedt. 2010. 'Evidence for natural antisense transcript-mediated inhibition of microRNA function', *Genome Biol*, 11: R56.

Fang, Z., R. Du, A. Edwards, E. K. Flemington, and K. Zhang. 2013. 'The sequence structures of human microRNA molecules and their implications', *PLoS One*, 8: e54215.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley, and J. A. Piccirilli. 2013. 'RNA catalyses nuclear pre-mRNA splicing', *Nature*, 503: 229-34.

Forsti, A., Q. Jin, A. Altieri, R. Johansson, K. Wagner, K. Enquist, E. Grzybowska, J. Pamula, W. Pekala, G. Hallmans, P. Lenner, and K. Hemminki. 2007. 'Polymorphisms in the KDR and POSTN genes: association with breast cancer susceptibility and prognosis', *Breast Cancer Res Treat*, 101: 83-93.

Frith, M. C., M. Pheasant, and J. S. Mattick. 2005. 'The amazing complexity of the human transcriptome', *Eur J Hum Genet*, 13: 894-7.

Fuks, F., W. A. Burgers, N. Godin, M. Kasai, and T. Kouzarides. 2001. 'Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription', *EMBO J*, 20: 2536-44.

Galetzka, D., E. Weis, T. Tralau, L. Seidmann, and T. Haaf. 2007. 'Sex-specific windows for high mRNA expression of DNA methyltransferases 1 and 3A and methyl-CpG-binding domain proteins 2 and 4 in human fetal gonads', *Mol Reprod Dev*, 74: 233-41.

Garcia, J. M., J. Silva, C. Pena, V. Garcia, R. Rodriguez, M. A. Cruz, B. Cantos, M. Provencio, P. Espana, and F. Bonilla. 2004. 'Promoter methylation of the PTEN gene is a common molecular change in breast cancer', *Genes Chromosomes Cancer*, 41: 117-24.

Gibb, E. A., C. J. Brown, and W. L. Lam. 2011. 'The functional role of long non-coding RNA in human carcinomas', *Mol Cancer*, 10: 38.

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

Goodstadt, L., and C. P. Ponting. 2006. 'Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human', *PLoS Comput Biol*, 2: e133.

Gritsenko, P. G., O. Ilina, and P. Friedl. 2012. 'Interstitial guidance of cancer invasion', *J Pathol*, 226: 185-99.

Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace, and S. Altman. 1983. 'The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme', *Cell*, 35: 849-57.

Guil, S., and M. Esteller. 2012. 'Cis-acting noncoding RNAs: friends and foes', *Nat Struct Mol Biol*, 19: 1068-75.

Gunia, S., A. Jain, S. Koch, S. Denzinger, S. Gotz, N. Niessl, and M. May. 2013. 'Periostin expression correlates with pT-stage, grading and tumour size, and independently predicts cancer-specific survival in surgically treated penile squamous cell carcinomas', *J Clin Pathol*.

Guo, H., N. T. Ingolia, J. S. Weissman, and D. P. Bartel. 2010. 'Mammalian microRNAs predominantly act to decrease target mRNA levels', *Nature*, 466: 835-40.

Gupta, R. A., N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M. C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar, and H. Y. Chang. 2010. 'Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis', *Nature*, 464: 1071-6.

Gutschner, T., and S. Diederichs. 2012. 'The hallmarks of cancer: a long non-coding RNA point of view', *RNA Biol*, 9: 703-19.

Guttman, M., I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn, and E. S. Lander. 2009. 'Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals', *Nature*, 458: 223-7.

Hamilton, D. W. 2008. 'Functional role of periostin in development and wound repair: implications for connective tissue disease', *J Cell Commun Signal*, 2: 9-17.

Han, J., D. Kim, and K. V. Morris. 2007. 'Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells', *Proc Natl Acad Sci U S A*, 104: 12422-7.

Hashimoto, H., S. Hong, A. S. Bhagwat, X. Zhang, and X. Cheng. 2012. 'Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation', *Nucleic Acids Res*, 40: 10203-14.

Hashimoto, H., Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang, and X. Cheng. 2012. 'Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation', *Nucleic Acids Res*, 40: 4841-9.

Hawkins, P. G., and K. V. Morris. 2010a. 'Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5', *Transcription*, 1: 165-75.

———. 2010b. 'Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5', *Transcr*, 1: 165-75.

Hawkins, P. G., S. Santoso, C. Adams, V. Anest, and K. V. Morris. 2009a. 'Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells', *Nucleic Acids Res*, 37: 2984-95.

———. 2009b. 'Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells', *Nucleic Acids Res*.

He, J., S. A. Whelan, M. Lu, D. Shen, D. U. Chung, R. E. Saxton, K. F. Faull, J. P. Whitelegge, and H. R. Chang. 2011. 'Proteomic-based biosignatures in breast cancer classification and prediction of therapeutic response', *Int J Proteomics*, 2011: 896476.

Hoagland, M. B., M. L. Stephenson, J. F. Scott, L. I. Hecht, and P. C. Zamecnik. 1958. 'A soluble ribonucleic acid intermediate in protein synthesis', *J Biol Chem*, 231: 241-57.

Hsu, M. T., and M. Coca-Prados. 1979. 'Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells', *Nature*, 280: 339-40.

Huang, R., and E. K. Rofstad. 2017. 'Cancer stem cells (CSCs), cervical CSCs and targeted therapies', *Oncotarget*, 8: 35351-67.

International Chicken Genome Sequencing, Consortium. 2004. 'Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution', *Nature*, 432: 695-716.

Jacobs, A. L., and P. Schar. 2012. 'DNA glycosylases: in DNA repair and beyond', *Chromosoma*, 121: 1-20.

Jacq, C., J. R. Miller, and G. G. Brownlee. 1977. 'A pseudogene structure in 5S DNA of Xenopus laevis', *Cell*, 12: 109-20.

Janowski, B. A., K. E. Huffman, J. C. Schwartz, R. Ram, D. Hardy, D. S. Shames, J. D. Minna, and D. R. Corey. 2005. 'Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs', *Nat Chem Biol*, 1: 216-22.

Janowski, B. A., K. E. Huffman, J. C. Schwartz, R. Ram, R. Nordsell, D. S. Shames, J. D. Minna, and D. R. Corey. 2006. 'Involvement of AGO1 and AGO2 in mammalian transcriptional silencing', *Nat Struct Mol Biol*, 13: 787-92.

Johnsson, P., A. Ackley, L. Vidarsdottir, W. O. Lui, M. Corcoran, D. Grander, and K. V. Morris. 2013a. 'A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells', *Nat Struct Mol Biol*.

———. 2013b. 'A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells', *Nat Struct Mol Biol*, 20: 440-6.

Johnsson, P., L. Lipovich, D. Grander, and K. V. Morris. 2013. 'Evolutionary conservation of long noncoding RNAs; sequence, structure, function', *Biochim Biophys Acta*.

———. 2014. 'Evolutionary conservation of long non-coding RNAs; sequence, structure, function', *Biochim Biophys Acta*, 1840: 1063-71.

Kanno, A., K. Satoh, A. Masamune, M. Hirota, K. Kimura, J. Umino, S. Hamada, A. Satoh, S. Egawa, F. Motoi, M. Unno, and T. Shimosegawa. 2008. 'Periostin, secreted from stromal cells, has biphasic effect on cell migration and correlates with the epithelial to mesenchymal transition of human pancreatic cancer cells', *Int J Cancer*, 122: 2707-18.

Katayama, S., Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, H. Nishida, C. C. Yap, M. Suzuki, J. Kawai, H. Suzuki, P. Carninci, Y. Hayashizaki, C. Wells, M. Frith, T. Ravasi, K. C. Pang, J. Hallinan, J. Mattick, D. A. Hume, L. Lipovich, S. Batalov, P. G. Engstrom, Y. Mizuno, M. A. Faghihi, A. Sandelin, A. M. Chalk, S. Mottagui-Tabar, Z. Liang, B. Lenhard, C. Wahlestedt, Riken Genome

Exploration Research Group, Group Genome Science, and Fantom Consortium. 2005. 'Antisense transcription in the mammalian transcriptome', *Science*, 309: 1564-6.

Kato, H., D. B. Goto, R. A. Martienssen, T. Urano, K. Furukawa, and Y. Murakami. 2005. 'RNA polymerase II is required for RNAi-dependent heterochromatin assembly', *Science*, 309: 467-9.

Kemmerich, K., F. A. Dingler, C. Rada, and M. S. Neuberger. 2012. 'Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice', *Nucleic Acids Res*, 40: 6016-25.

Khachane, A. N., and P. M. Harrison. 2009. 'Assessing the genomic evidence for conserved transcribed pseudogenes under selection', *BMC Genomics*, 10: 435.

Khalil, A. M., M. Guttman, M. Huarte, M. Garber, A. Raj, D. Rivea Morales, K. Thomas, A. Presser, B. E. Bernstein, A. van Oudenaarden, A. Regev, E. S. Lander, and J. L. Rinn. 2009. 'Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression', *Proc Natl Acad Sci U S A*, 106: 11667-72.

Kharaishvili, G., M. Cizkova, K. Bouchalova, G. Mgebrishvili, Z. Kolar, and J. Bouchal. 2011. 'Collagen triple helix repeat containing 1 protein, periostin and versican in primary and metastatic breast cancer: an immunohistochemical study', *J Clin Pathol*, 64: 977-82.

Kikuchi, Y., T. G. Kashima, T. Nishiyama, K. Shimazu, Y. Morishita, M. Shimazaki, I. Kii, H. Horie, H. Nagai, A. Kudo, and M. Fukayama. 2008. 'Periostin is expressed in pericryptal fibroblasts and cancer-associated fibroblasts in the colon', *J Histochem Cytochem*, 56: 753-64.

Kim, C. J., N. Yoshioka, Y. Tambe, R. Kushima, Y. Okada, and H. Inoue. 2005. 'Periostin is down-regulated in high grade human bladder cancers and suppresses in vitro cell invasiveness and in vivo metastasis of cancer cells', *Int J Cancer*, 117: 51-8.

Kim, D. H., P. Saetrom, O. Snove, Jr., and J. J. Rossi. 2008. 'MicroRNA-directed transcriptional gene silencing in mammalian cells', *Proc Natl Acad Sci U S A*, 105: 16230-5.

Kim, J. W., Y. H. Zhang, M. A. Zern, J. J. Rossi, and J. Wu. 2007. 'Short hairpin RNA causes the methylation of transforming growth factor-beta receptor II promoter and silencing of the target gene in rat hepatic stellate cells', *Biochem Biophys Res Commun*, 359: 292-7.

Kiss, T. 2001. 'Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs', *EMBO J*, 20: 3617-22.

Knowling, S., K. Stapleton, A. M. Turner, E. Uhlmann, T. Lehmann, J. Vollmer, and K. V. Morris. 2012. 'Chemically Modified Oligonucleotides Modulate an Epigenetically Varied and Transient Form of Transcription Silencing of HIV-1 in Human Cells', *Mol Ther Nucleic Acids*, 1: e16.

Kondo, E., Z. Gu, A. Horii, and S. Fukushige. 2005. 'The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes', *Mol Cell Biol*, 25: 4388-96.

Kristensen, L. S., T. B. Hansen, M. T. Veno, and J. Kjems. 2018. 'Circular RNAs in cancer: opportunities and challenges in the field', *Oncogene*, 37: 555-65.

Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech. 1982. 'Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena', *Cell*, 31: 147-57.

Kudo, Y., S. Iizuka, M. Yoshida, P. T. Nguyen, S. B. Siriwardena, T. Tsunematsu, M. Ohbayashi, T. Ando, D. Hatakeyama, T. Shibata, K. Koizumi, M. Maeda, N. Ishimaru, I. Ogawa, and T. Takata. 2012. 'Periostin directly and indirectly promotes tumor lymphangiogenesis of head and neck cancer', *PLoS One*, 7: e44488.

Kudo, Y., I. Ogawa, S. Kitajima, M. Kitagawa, H. Kawai, P. M. Gaffney, M. Miyauchi, and T. Takata. 2006. 'Periostin promotes invasion and anchorage-independent growth in the metastatic process of head and neck cancer', *Cancer Res*, 66: 6928-35.

Kudo, Y., B. S. Siriwardena, H. Hatano, I. Ogawa, and T. Takata. 2007. 'Periostin: novel diagnostic and therapeutic target for cancer', *Histol Histopathol*, 22: 1167-74.

Kunz, C., F. Focke, Y. Saito, D. Schuermann, T. Lettieri, J. Selfridge, and P. Schar. 2009. 'Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil', *PLoS Biol*, 7: e91.

Kyutoku, M., Y. Taniyama, N. Katsuragi, H. Shimizu, Y. Kunugiza, K. Iekushi, N. Koibuchi, F. Sanada, Y. Oshita, and R. Morishita. 2011. 'Role of periostin in cancer progression and metastasis: inhibition of breast cancer progression and metastasis by anti-periostin antibody in a murine model', *Int J Mol Med*, 28: 181-6.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin,

I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki, and Consortium International Human Genome Sequencing. 2001. 'Initial sequencing and analysis of the human genome', *Nature*, 409: 860-921.

Lee, J. T., L. S. Davidow, and D. Warshawsky. 1999. 'Tsix, a gene antisense to Xist at the X-inactivation centre', *Nat Genet*, 21: 400-4.

Lee, R. C., R. L. Feinbaum, and V. Ambros. 1993. 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14', *Cell*, 75: 843-54.

Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V. N. Kim. 2003. 'The nuclear RNase III Drosha initiates microRNA processing', *Nature*, 425: 415-9.

Lerner, M., M. Harada, J. Loven, J. Castro, Z. Davis, D. Oscier, M. Henriksson, O. Sangfelt, D. Grander, and M. M. Corcoran. 2009. 'DLEU2, frequently deleted in malignancy, functions as a critical host gene of the cell cycle inhibitory microRNAs miR-15a and miR-16-1', *Exp Cell Res*, 315: 2941-52.

Lewis, B. P., C. B. Burge, and D. P. Bartel. 2005. 'Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets', *Cell*, 120: 15-20.

Li, B., L. Wang, and B. Chi. 2013. 'Upregulation of periostin prevents P53-mediated apoptosis in SGC-7901 gastric cancer cells', *Mol Biol Rep*, 40: 1677-83.

Li, J. S., G. W. Sun, X. Y. Wei, and W. H. Tang. 2007. 'Expression of periostin and its clinicopathological relevance in gastric cancer', *World J Gastroenterol*, 13: 5261-6.

Lim, H. G., K. Suzuki, D. A. Cooper, and A. D. Kelleher. 2008. 'Promoter-targeted siRNAs induce gene silencing of simian immunodeficiency virus (SIV) infection in vitro', *Mol Ther*, 16: 565-70.

Lippman, Z., and R. Martienssen. 2004. 'The role of RNA interference in heterochromatic silencing', *Nature*, 431: 364-70.

Lister, N., G. Shevchenko, J. L. Walshe, J. Groen, P. Johnsson, L. Vidarsdottir, D. Grander, S. F. Ataide, and K. V. Morris. 2017. 'The molecular dynamics of long noncoding RNA control of transcription in PTEN and its pseudogene', *Proc Natl Acad Sci U S A*, 114: 9942-47.

Lister, Nicholas C., Matthew Clemson, and Kevin V. Morris. 2015. 'RNA-directed epigenetic silencing of Periostin inhibits cell motility', *Royal Society Open Science*, 2.

Lister, R., R. C. O'Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar, and J. R. Ecker. 2008. 'Highly integrated single-base resolution maps of the epigenome in Arabidopsis', *Cell*, 133: 523-36.

Liu, Y., and B. A. Liu. 2011. 'Enhanced proliferation, invasion, and epithelial-mesenchymal transition of nicotine-promoted gastric cancer by periostin', *World J Gastroenterol*, 17: 2674-80.

Louro, R., T. El-Jundi, H. I. Nakaya, E. M. Reis, and S. Verjovski-Almeida. 2008. 'Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci', *Genomics*, 92: 18-25.

Lutsenko, E., and A. S. Bhagwat. 1999. 'Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. A model, its experimental support and implications', *Mutat Res*, 437: 11-20.

Mahmoudi, S., S. Henriksson, M. Corcoran, C. Mendez-Vidal, K. G. Wiman, and M. Farnebo. 2009. 'Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage', *Mol Cell*, 33: 462-71.

Malanchi, I., and J. Huelsken. 2009. 'Cancer stem cells: never Wnt away from the niche', *Curr Opin Oncol*, 21: 41-6.

Malanchi, I., A. Santamaria-Martinez, E. Susanto, H. Peng, H. A. Lehr, J. F. Delaloye, and J. Huelsken. 2012. 'Interactions between cancer stem cells and their niche govern metastatic colonization', *Nature*, 481: 85-9.

Mattick, J. S., and I. V. Makunin. 2006. 'Non-coding RNA', *Hum Mol Genet*, 15 Spec No 1: R17-29.

Matzke, M. A., and J. A. Birchler. 2005. 'RNAi-mediated pathways in the nucleus', *Nat Rev Genet*, 6: 24-35.

Matzke, M. A., and R. A. Mosher. 2014. 'RNA-directed DNA methylation: an epigenetic pathway of increasing complexity', *Nat Rev Genet*, 15: 394-408.

Matzke, M. A., M. Primig, J. Trnovsky, and A. J. Matzke. 1989. 'Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants', *EMBO J*, 8: 643-9.

Mayo, L. D., and D. B. Donner. 2001. 'A phosphatidylinositol 3-kinase/Akt pathway promotes translocation of Mdm2 from the cytoplasm to the nucleus', *Proc Natl Acad Sci U S A*, 98: 11598-603.

Memczak, S., M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. le Noble, and N. Rajewsky. 2013. 'Circular RNAs are a large class of animal RNAs with regulatory potency', *Nature*, 495: 333-8.

Meng, F., R. Henson, H. Wehbe-Janek, K. Ghoshal, S. T. Jacob, and T. Patel. 2007. 'MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer', *Gastroenterology*, 133: 647-58.

Michaylira, C. Z., G. S. Wong, C. G. Miller, C. M. Gutierrez, H. Nakagawa, R. Hammond, A. J. Klein-Szanto, J. S. Lee, S. B. Kim, M. Herlyn, J. A. Diehl, P. Gimotty, and A. K. Rustgi. 2010. 'Periostin, a cell adhesion molecule, facilitates invasion in the tumor microenvironment and annotates a novel tumor-invasive signature in esophageal cancer', *Cancer Res*, 70: 5281-92.

Mirmohammadsadegh, A., A. Marini, S. Nambiar, M. Hassan, A. Tannapfel, T. Ruzicka, and U. R. Hengge. 2006. 'Epigenetic silencing of the PTEN gene in melanoma', *Cancer Res*, 66: 6546-52.

Morra, L., and H. Moch. 2011. 'Periostin expression and epithelial-mesenchymal transition in cancer: a review and an update', *Virchows Arch*, 459: 465-75.

Morra, L., M. Rechsteiner, S. Casagrande, V. Duc Luu, R. Santimaria, P. A. Diener, T. Sulser, G. Kristiansen, P. Schraml, H. Moch, and A. Soltermann. 2011. 'Relevance of periostin splice variants in renal cell carcinoma', *Am J Pathol*, 179: 1513-21.

Morra, L., M. Rechsteiner, S. Casagrande, A. von Teichman, P. Schraml, H. Moch, and A. Soltermann. 2012. 'Characterization of periostin isoform pattern in non-small cell lung cancer', *Lung Cancer*, 76: 183-90.

Morris, K. V. 2009a. 'Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells', *Epigenetics*, 4.

———. 2009b. 'Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells', *Epigenetics*, 4: 296-301.

———. 2009c. 'RNA-directed transcriptional gene silencing and activation in human cells', *Oligonucleotides*, 19: 299-306.

———. 2015. 'The theory of RNA-mediated gene evolution', *Epigenetics*, 10: 1-5.

Morris, K. V., S. W. Chan, S. E. Jacobsen, and D. J. Looney. 2004. 'Small interfering RNA-induced transcriptional gene silencing in human cells', *Science*, 305: 1289-92.

Morris, K. V., and J. S. Mattick. 2014a. 'The rise of regulatory RNA', *Nat Rev Genet*, 15: 423-37.

———. 2014b. 'The rise of regulatory RNA', *Nat Rev Genet*.

Morris, K. V., S. Santoso, A. M. Turner, C. Pastori, and P. G. Hawkins. 2008. 'Bidirectional transcription directs both transcriptional gene activation and suppression in human cells', *PLoS Genet*, 4: e1000258.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. 'Mapping and quantifying mammalian transcriptomes by RNA-Seq', *Nat Methods*, 5: 621-8.

Mouse Genome Sequencing, Consortium, R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody, and E. S. Lander. 2002. 'Initial sequencing and comparative analysis of the mouse genome', *Nature*, 420: 520-62.

Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder. 2008. 'The transcriptional landscape of the yeast genome defined by RNA sequencing', *Science*, 320: 1344-9.

Napoli, C., C. Lemieux, and R. Jorgensen. 1990. 'Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans', *Plant Cell*, 2: 279-89.

Napoli, S., C. Pastori, M. Magistri, G. M. Carbone, and C. V. Catapano. 2009a. 'Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells', *Embo J*.

———. 2009b. 'Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells', *EMBO J*, 28: 1708-19.

Neddermann, P., P. Gallinari, T. Lettieri, D. Schmid, O. Truong, J. J. Hsuan, K. Wiebauer, and J. Jiricny. 1996. 'Cloning and expression of human G/T mismatch-specific thymine-DNA glycosylase', *J Biol Chem*, 271: 12767-74.

Nuzzo, P. V., A. Rubagotti, L. Zinoli, F. Ricci, S. Salvi, S. Boccardo, and F. Boccardo. 2012. 'Prognostic value of stromal and epithelial periostin expression in human prostate cancer: correlation with clinical pathological features and the risk of biochemical relapse or death', *BMC Cancer*, 12: 625.

Ohno, S. 1972. 'So much "junk" DNA in our genome', *Brookhaven Symp Biol*, 23: 366-70.

Ohno, S., and T. Yomo. 1991. 'The grammatical rule for all DNA: junk and coding sequences', *Electrophoresis*, 12: 103-8.

Ohshima, K., M. Hattori, T. Yada, T. Gojobori, Y. Sakaki, and N. Okada. 2003. 'Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates', *Genome Biol*, 4: R74.

Omoto, S., and Y. R. Fujii. 2005. 'Regulation of human immunodeficiency virus 1 transcription by nef microRNA', *J Gen Virol*, 86: 751-5.

Omoto, S., M. Ito, Y. Tsutsumi, Y. Ichikawa, H. Okuyama, E. A. Brisibe, N. K. Saksena, and Y. R. Fujii. 2004. 'HIV-1 nef suppression by virally encoded microRNA', *Retrovirology*, 1: 44.

Ouyang, G., M. Liu, K. Ruan, G. Song, Y. Mao, and S. Bao. 2009. 'Upregulated expression of periostin by hypoxia in non-small-cell lung cancer cells promotes cell survival via the Akt/PKB pathway', *Cancer Lett*, 281: 213-9.

Palade, G. E. 1955. 'A small particulate component of the cytoplasm', *J Biophys Biochem Cytol*, 1: 59-68.

Pei, B., C. Sisu, A. Frankish, C. Howald, L. Habegger, X. J. Mu, R. Harte, S. Balasubramanian, A. Tanzer, M. Diekhans, A. Reymond, T. J. Hubbard, J. Harrow, and M. B. Gerstein. 2012. 'The GENCODE pseudogene resource', *Genome Biol*, 13: R51.

Pennisi, E. 2012. 'Genomics. ENCODE project writes eulogy for junk DNA', *Science*, 337: 1159-61.

Peters, L., and G. Meister. 2007. 'Argonaute proteins: mediators of RNA silencing', *Mol Cell*, 26: 611-23.

Poliseno, L., A. Haimovic, P. J. Christos, Y. Saenz de Miera E. C. Vega, R. Shapiro, A. Pavlick, R. S. Berman, F. Darvishian, and I. Osman. 2011. 'Deletion of PTENP1 pseudogene in human melanoma', *J Invest Dermatol*, 131: 2497-500.

Poliseno, L., and P. P. Pandolfi. 2015. 'PTEN ceRNA networks in human cancer', *Methods*, 77-78: 41-50.

Poliseno, L., L. Salmena, J. Zhang, B. Carver, W. J. Haveman, and P. P. Pandolfi. 2010. 'A coding-independent function of gene and pseudogene mRNAs regulates tumour biology', *Nature*, 465: 1033-8.

Prensner, J. R., and A. M. Chinnaiyan. 2011. 'The emergence of lncRNAs in cancer biology', *Cancer Discov*, 1: 391-407.

Puget, N., S. Gad, L. Perrin-Vidoz, O. M. Sinilnikova, D. Stoppa-Lyonnet, G. M. Lenoir, and S. Mazoyer. 2002. 'Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot', *Am J Hum Genet*, 70: 858-65.

Puglisi, F., C. Puppin, E. Pegolo, C. Andreetta, G. Pascoletti, F. D'Aurizio, M. Pandolfi, G. Fasola, A. Piga, G. Damante, and C. Di Loreto. 2008. 'Expression of periostin in human breast cancer', *J Clin Pathol*, 61: 494-8.

Rahman, L., V. Bliskovski, F. J. Kaye, and M. Zajac-Kaye. 2004. 'Evolutionary conservation of a 2-kb intronic sequence flanking a tissue-specific alternative exon in the PTBP2 gene', *Genomics*, 83: 76-84.

Ramachandran, P. V., and S. Ignacimuthu. 2013. 'RNA interference--a silent but an efficient therapeutic tool', *Appl Biochem Biotechnol*, 169: 1774-89.

Rinn, J. L., and H. Y. Chang. 2012. 'Genome regulation by long noncoding RNAs', *Annu Rev Biochem*, 81: 145-66.

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, and H. Y. Chang. 2007. 'Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs', *Cell*, 129: 1311-23.

Roberts, T. C., S. E. Andaloussi, K. V. Morris, G. McClorey, and M. J. Wood. 2012. 'Small RNA-Mediated Epigenetic Myostatin Silencing', *Mol Ther Nucleic Acids*, 1: e23.

Roest Crollius, H., O. Jaillon, A. Bernot, C. Dasilva, L. Bouneau, C. Fischer, C. Fizames, P. Wincker, P. Brottier, F. Quetier, W. Saurin, and J. Weissenbach. 2000. 'Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence', *Nat Genet*, 25: 235-8.

Rosenbloom, K. R., T. R. Dreszer, J. C. Long, V. S. Malladi, C. A. Sloan, B. J. Raney, M. S. Cline, D. Karolchik, G. P. Barber, H. Clawson, M. Diekhans, P. A. Fujita, M. Goldman, R. C. Gravell, R. A. Harte, A. S. Hinrichs, V. M. Kirkup, R. M. Kuhn, K. Learned, M. Maddren, L. R. Meyer, A. Pohl, B. Rhead, M. C. Wong, A. S. Zweig, D. Haussler, and W. J. Kent. 2012. 'ENCODE whole-genome data in the UCSC Genome Browser: update 2012', *Nucleic Acids Res*, 40: D912-7.

Ruan, K., S. Bao, and G. Ouyang. 2009. 'The multifaceted role of periostin in tumorigenesis', *Cell Mol Life Sci*, 66: 2219-30.

Saal, L. H., S. K. Gruvberger-Saal, C. Persson, K. Lovgren, M. Jumppanen, J. Staaf, G. Jonsson, M. M. Pires, M. Maurer, K. Holm, S. Koujak, S. Subramaniyam, J. Vallon-Christersson, H. Olsson, T. Su, L. Memeo, T. Ludwig, S. P. Ethier, M. Krogh, M. Szabolcs, V. V. Murty, J. Isola, H. Hibshoosh, R. Parsons, and A. Borg. 2008. 'Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair', *Nat Genet*, 40: 102-7.

Saayman, S., A. Ackley, A. M. Turner, M. Famiglietti, A. Bosque, M. Clemson, V. Planelles, and K. V. Morris. 2014a. 'An HIV-Encoded Antisense Long Noncoding RNA Epigenetically Regulates Viral Transcription', *Mol Ther*, 22: 1164-75.

Saayman, S., A. Ackley, A. W. Turner, M. Famiglietti, A. Bosque, M. Clemson, V. Planelles, and K. V. Morris. 2014b. 'An HIV-encoded antisense long noncoding RNA epigenetically regulates viral transcription', *Mol Ther*, 22: 1164-75.

Salmena, L., A. Carracedo, and P. P. Pandolfi. 2008. 'Tenets of PTEN tumor suppression', *Cell*, 133: 403-14.

Sasaki, H., C. Y. Yu, M. Dai, C. Tam, M. Loda, D. Auclair, L. B. Chen, and A. Elias. 2003. 'Elevated serum periostin levels in patients with bone metastases from breast but not lung cancer', *Breast Cancer Res Treat*, 77: 245-52.

Scarola, M., E. Comisso, R. Pascolo, R. Chiaradia, R. Maria Marion, C. Schneider, M. A. Blasco, S. Schoeftner, and R. Benetti. 2015. 'Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA', *Nat Commun*, 6: 7631.

Scarola, M., E. Comisso, R. Pascolo, R. Chiaradia, R. M. Marion, C. Schneider, M. A. Blasco, S. Schoeftner, and R. Benetti. 2015. 'Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA', *Nat Commun*, 6: 7631.

Shen, J. C., W. M. Rideout, 3rd, and P. A. Jones. 1994. 'The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA', *Nucleic Acids Res*, 22: 972-6.

Singh, A., J. K. Palanichamy, P. Ramalingam, M. A. Kassab, M. Bhagat, R. Andrabi, K. Luthra, S. Sinha, and P. Chattopadhyay. 2014. 'Long-term suppression of HIV-1C virus production in human peripheral blood mononuclear cells by LTR heterochromatization with a short double-stranded RNA', *J Antimicrob Chemother*, 69: 404-15.

Siriwardena, B. S., Y. Kudo, I. Ogawa, M. Kitagawa, S. Kitajima, H. Hatano, W. M. Tilakaratne, M. Miyauchi, and T. Takata. 2006. 'Periostin is frequently overexpressed and enhances invasion and angiogenesis in oral cancer', *Br J Cancer*, 95: 1396-403.

Sjolund, A. B., A. G. Senejani, and J. B. Sweasy. 2013. 'MBD4 and TDG: multifaceted DNA glycosylases with ever expanding biological roles', *Mutat Res*, 743-744: 12-25.

Soltermann, A. 2012. '[Epithelial-mesenchymal transition in non-small cell lung cancer]', *Pathologe*, 33 Suppl 2: 311-7.

Soltermann, A., V. Tischler, S. Arbogast, J. Braun, N. Probst-Hensch, W. Weder, H. Moch, and G. Kristiansen. 2008. 'Prognostic significance of epithelial-mesenchymal and mesenchymal-epithelial transition protein expression in non-small cell lung cancer', *Clin Cancer Res*, 14: 7430-7.

Stambolic, V., D. MacPherson, D. Sas, Y. Lin, B. Snow, Y. Jang, S. Benchimol, and T. W. Mak. 2001. 'Regulation of PTEN transcription by p53', *Mol Cell*, 8: 317-25.

Steck, P. A., M. A. Pershouse, S. A. Jasser, W. K. Yung, H. Lin, A. H. Ligon, L. A. Langford, M. L. Baumgard, T. Hattier, T. Davis, C. Frye, R. Hu, B. Swedlund, D. H. Teng, and S. V. Tavtigian. 1997. 'Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers', *Nat Genet*, 15: 356-62.

Streit, S., C. W. Michalski, M. Erkan, H. Friess, and J. Kleeff. 2009. 'Confirmation of DNA microarray-derived differentially expressed genes in pancreatic cancer using quantitative RT-PCR', *Pancreatology*, 9: 577-82.

Sun, C., C. Song, Z. Ma, K. Xu, Y. Zhang, H. Jin, S. Tong, W. Ding, G. Xia, and Q. Ding. 2011. 'Periostin identified as a potential biomarker of prostate cancer by iTRAQ-proteomics analysis of prostate biopsy', *Proteome Sci*, 9: 22.

Sun, C., X. Zhao, K. Xu, J. Gong, W. Liu, W. Ding, Y. Gou, G. Xia, and Q. Ding. 2011. 'Periostin: a promising target of therapeutical intervention for prostate cancer', *J Transl Med*, 9: 99.

Suo, G., J. Han, X. Wang, J. Zhang, Y. Zhao, Y. Zhao, and J. Dai. 2005. 'Oct4 pseudogenes are transcribed in cancers', *Biochem Biophys Res Commun*, 337: 1047-51.

Suspene, R., M. M. Aynaud, J. P. Vartanian, and S. Wain-Hobson. 2013. 'Efficient deamination of 5-methylcytidine and 5-substituted cytidine residues in DNA by human APOBEC3A cytidine deaminase', *PLoS One*, 8: e63461.

Suzuki, K., S. Hattori, K. Marks, C. Ahlenstiel, Y. Maeda, T. Ishida, M. Millington, M. Boyd, G. Symonds, D. A. Cooper, S. Okada, and A. D. Kelleher. 2013. 'Promoter Targeting shRNA Suppresses HIV-1 Infection In vivo Through Transcriptional Gene Silencing', *Mol Ther Nucleic Acids*, 2: e137.

Suzuki, K., T. Ishida, M. Yamagishi, C. Ahlenstiel, S. Swaminathan, K. Marks, D. Murray, E. M. McCartney, M. R. Beard, M. Alexander, D. F. Purcell, D. A. Cooper, T. Watanabe, and A. D. Kelleher. 2011. 'Transcriptional gene silencing of HIV-1 through promoter targeted RNA is highly specific', *RNA Biol*, 8: 1035-46.

Suzuki, K., T. Juelich, H. Lim, T. Ishida, T. Watanebe, D. A. Cooper, S. Rao, and A. D. Kelleher. 2008. 'Closed chromatin architecture is induced by an RNA duplex targeting the HIV-1 promoter region', *J Biol Chem*, 283: 23353-63.

Taft, R. J., P. G. Hawkins, J. S. Mattick, and K. V. Morris. 2011. 'The relationship between transcription initiation RNAs and CCCTC-binding factor (CTCF) localization', *Epigenetics Chromatin*, 4: 13.

Taft, R. J., M. Pheasant, and J. S. Mattick. 2007. 'The relationship between non-protein-coding DNA and eukaryotic complexity', *Bioessays*, 29: 288-99.

Takanami, I., T. Abiko, and S. Koizumi. 2008. 'Expression of periostin in patients with non-small cell lung cancer: correlation with angiogenesis and lymphangiogenesis', *Int J Biol Markers*, 23: 182-6.

Takeshita, S., R. Kikuno, K. Tezuka, and E. Amann. 1993. 'Osteoblast-specific factor 2: cloning of a putative bone adhesion protein with homology with the insect protein fasciclin I', *Biochem J*, 294 (Pt 1): 271-8.

Teng, D. H., R. Hu, H. Lin, T. Davis, D. Iliev, C. Frye, B. Swedlund, K. L. Hansen, V. L. Vinson, K. L. Gumpper, L. Ellis, A. El-Naggar, M. Frazier, S. Jasser, L. A. Langford, J. Lee, G. B. Mills, M. A. Pershouse, R. E. Pollack, C. Tornos, P. Troncoso, W. K. Yung, G. Fujii, A. Berson, P. A. Steck, and et al. 1997. 'MMAC1/PTEN mutations in primary tumor specimens and tumor cell lines', *Cancer Res*, 57: 5221-5.

Tian, B., Y. Zhang, N. Li, X. Liu, and J. Dong. 2012. 'CD74: a potential novel target for triple-negative breast cancer', *Tumour Biol*, 33: 2273-7.

Tian, Y., G. S. Bova, and H. Zhang. 2011. 'Quantitative glycoproteomic analysis of optimal cutting temperature-embedded frozen tissues identifying glycoproteins associated with aggressive prostate cancer', *Anal Chem*, 83: 7013-9.

Tischler, V., F. R. Fritzsche, P. J. Wild, C. Stephan, H. H. Seifert, M. O. Riener, T. Hermanns, A. Mortezavi, J. Gerhardt, P. Schraml, K. Jung, H. Moch, A. Soltermann, and G. Kristiansen. 2010. 'Periostin is up-regulated in high grade and high stage prostate cancer', *BMC Cancer*, 10: 273.

Tricarico, R., S. Cortellino, A. Riccio, S. Jagmohan-Changur, H. Van der Klift, J. Wijnen, D. Turner, A. Ventura, V. Rovella, A. Percesepe, E. Lucci-Cordisco, P. Radice, L. Bertario, M. Pedroni, M. Ponz de Leon, P. Mancuso, K. Devarajan, K. Q. Cai, A. J. Klein-Szanto, G. Neri, P. Moller, A. Viel, M. Genuardi, R. Fodde, and A. Bellacosa. 2015. 'Involvement of MBD4 inactivation in mismatch repair-deficient tumorigenesis', *Oncotarget*, 6: 42892-904.

Trotman, L. C., M. Niki, Z. A. Dotan, J. A. Koutcher, A. Di Cristofano, A. Xiao, A. S. Khoo, P. Roy-Burman, N. M. Greenberg, T. Van Dyke, C. Cordon-Cardo, and P. P. Pandolfi. 2003. 'Pten dose dictates cancer progression in the prostate', *PLoS Biol*, 1: E59.

Tsai, M. C., O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal, and H. Y. Chang. 2010. 'Long noncoding RNA as modular scaffold of histone modification complexes', *Science*, 329: 689-93.

Tsunoda, T., B. Furusato, Y. Takashima, S. Ravulapalli, A. Dobi, S. Srivastava, D. G. McLeod, I. A. Sesterhenn, D. K. Ornstein, and S. Shirasawa. 2009. 'The increased expression of periostin during early stages of prostate cancer and advanced stages of cancer stroma', *Prostate*, 69: 1398-403.

Tumbarello, D. A., J. Temple, and J. D. Brenton. 2012. 'ss3 integrin modulates transforming growth factor beta induced (TGFBI) function and paclitaxel response in ovarian cancer cells', *Mol Cancer*, 11: 36.

Turner, A. M., A. M. Ackley, M. A. Matrone, and K. V. Morris. 2012. 'Characterization of an HIV-targeted transcriptional gene-silencing RNA in primary cells', *Hum Gene Ther*, 23: 473-83.

Turner, A. M., J. De La Cruz, and K. V. Morris. 2009. 'Mobilization-competent Lentiviral Vector-mediated Sustained Transcriptional Modulation of HIV-1 Expression', *Mol Ther*, 17: 360-8.

Turner, A. M., and K. V. Morris. 2010. 'Controlling transcription with noncoding RNAs in mammalian cells', *Biotechniques*, 48: ix-xvi.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J.

Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. 2001. 'The sequence of the human genome', *Science*, 291: 1304-51.

Walker, S. M., N. R. Leslie, N. M. Perera, I. H. Batty, and C. P. Downes. 2004. 'The tumour-suppressor function of PTEN requires an N-terminal lipid-binding motif', *Biochem J*, 379: 301-7.

Wang, Z., and G. Ouyang. 2012. 'Periostin: a bridge between cancer stem cells and their metastatic niche', *Cell Stem Cell*, 10: 111-2.

Wassenegger, M., S. Heimes, L. Riedel, and H. L. Sanger. 1994. 'RNA-directed de novo methylation of genomic sequences in plants', *Cell*, 76: 567-76.

Waterhouse, P. M., M. W. Graham, and M. B. Wang. 1998. 'Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA', *Proc Natl Acad Sci U S A*, 95: 13959-64.

Wei, B., Y. Wei, K. Zhang, J. Wang, R. Xu, S. Zhan, G. Lin, W. Wang, M. Liu, L. Wang, R. Zhang, and J. Li. 2009. 'Development of an antisense RNA delivery system using conjugates of the MS2 bacteriophage capsids and HIV-1 TAT cell-penetrating peptide', *Biomed Pharmacother*, 63: 313-8.

Wei, J. X., J. Yang, J. F. Sun, L. T. Jia, Y. Zhang, H. Z. Zhang, X. Li, Y. L. Meng, L. B. Yao, and A. G. Yang. 2009. 'Both strands of siRNA have potential to guide posttranscriptional gene silencing in mammalian cells', *PLoS One*, 4: e5382.

Weinberg, M. S., and K. V. Morris. 2016. 'Transcriptional gene silencing in humans', *Nucleic Acids Res*, 44: 6505-17.

Weinberg, M. S., L. M. Villeneuve, A. Ehsani, M. Amarzguioui, L. Aagaard, Z. X. Chen, A. D. Riggs, J. J. Rossi, and K. V. Morris. 2006. 'The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells', *RNA*, 12: 256-62.

Wijesinghe, P., and A. S. Bhagwat. 2012. 'Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G', *Nucleic Acids Res*, 40: 9206-17.

Wu, G., X. Wang, and X. Zhang. 2013. 'Clinical Implications of Periostin in the Liver Metastasis of Colorectal Cancer', *Cancer Biother Radiopharm*.

Xing, M. 2010. 'Genetic alterations in the phosphatidylinositol-3 kinase/Akt pathway in thyroid cancer', *Thyroid*, 20: 697-706.

Xu, D., H. Xu, Y. Ren, C. Liu, X. Wang, H. Zhang, and P. Lu. 2012. 'Cancer stem cell-related gene periostin: a novel prognostic marker for breast cancer', *PLoS One*, 7: e46670.

Yap, K. L., S. Li, A. M. Munoz-Cabello, S. Raguz, L. Zeng, S. Mujtaba, J. Gil, M. J. Walsh, and M. M. Zhou. 2010. 'Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a', *Mol Cell*, 38: 662-74.

Younger, S. T., and D. R. Corey. 2011. 'Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters', *Nucleic Acids Res*, 39: 5682-91.

Yu, W., D. Gius, P. Onyango, K. Muldoon-Jacobs, J. Karp, A. P. Feinberg, and H. Cui. 2008. 'Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA', *Nature*, 451: 202-6.

Zemojtel, T., S. M. Kielbasa, P. F. Arndt, S. Behrens, G. Bourque, and M. Vingron. 2011. 'CpG deamination creates transcription factor-binding sites with high efficiency', *Genome Biol Evol*, 3: 1304-11.

Zeng, Y., R. Yi, and B. R. Cullen. 2003. 'MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms', *Proc Natl Acad Sci U S A*, 100: 9779-84.

Zhang, Y., and B. Tycko. 1992. 'Monoallelic expression of the human H19 gene', *Nat Genet*, 1: 40-4.

Zhang, Y., G. Zhang, J. Li, Q. Tao, and W. Tang. 2010. 'The expression analysis of periostin in human breast cancer', *J Surg Res*, 160: 102-6.

7 Appendices – Published Manuscripts

Chapter contents:

Page No.

7.1 Paper I

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

Cite this article: Lister NC, Clemson M, Morris KV. 2015 RNA-directed epigenetic silencing of Periostin inhibits cell motility. *R. Soc. open sci.* **2**: 140545. http://dx.doi.org/10.1098/rsos.140545

Received: 26 December 2014
Accepted: 14 May 2015

Subject Category:
Cellular and molecular biology

Subject Areas:
molecular biology/genetics/biochemistry

Keywords:
Periostin, metastasis, RNA, transcription, gene silencing, epigenetic

Author for correspondence:
Kevin V. Morris
e-mail: kmorris@unsw.edu.au

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsos.140545 or via http://rsos.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

RNA-directed epigenetic silencing of Periostin inhibits cell motility

Nicholas C. Lister[1], Matthew Clemson[1] and Kevin V. Morris[1,2]

[1]School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales, Australia
[2]Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA

The over-expression of Periostin, a member of the fasciclin family of proteins, has been reported in a number of cancers and, in particular, in metastatic tumours. These include breast, ovarian, lung, colon, head and neck, pancreatic, prostate, neuroblastoma and thyroid cancers. It is thought that Periostin plays a major role in the development of metastases owing to its apparent involvement in restructuring of the extracellular matrix to create a microenvironment favouring invasion and metastases, angiogenesis, independent proliferation, avoidance of apoptosis and the ability for cells to re-enter the cell cycle. As such we reasoned that targeted suppression of Periostin at the promoter and epigenetic level could result in the stable inhibition of cell motility. We find here that promoter-directed small antisense non-coding RNAs can induce transcriptional gene silencing of Periostin that results ultimately in a loss of cellular motility. The observations presented here suggest that cell motility and possibly metastasis can be controlled by transcriptional and epigenetic regulation of Periostin, offering a potentially new and novel manner to control the spread of cancerous cells.

1. Introduction

Periostin is an approximately 90 kDa matricellular protein capable of altering interactions between the extracellular matrix and surrounding cells [1,2]. Its structure contains a fourfold repeating domain that shows sequence homology with fasciclin 1, a protein found in insects, which interacts with integrins at the plasma membrane to accommodate motility and cell adhesion [1,3,4]. Binding of Periostin to various integrins activates the AKT/PKB and FAK-mediated signalling pathway, which promotes angiogenesis and cell survival, and is also involved in tumourigenesis by enabling enhanced tumour cell survival and proliferative and metastatic ability [1,5–7].

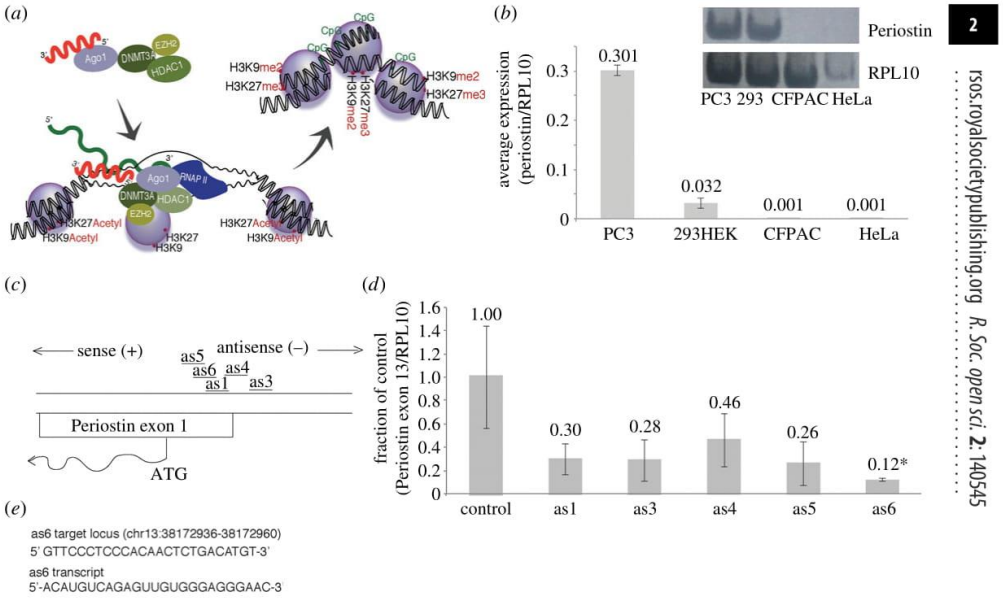


Figure 1. Characterization of Periostin expression and knockdown. (*a*) Schematic depicting sasRNA-directed TGS. The sasRNAs interact with Argonaute 1 (AGO1), DNA methyltransferase 3a (DNMT3a), histone deacetylase 1 (HDAC1) and Enhanzer of zeste 2 (EZH2) to epigenetically remodel target loci resulting in chromatin compaction and transcriptional silencing. (*b*) Quantitative RT-PCR (qRT-PCR) assay of endogenous expression of Periostin transcripts in various cell lines, normalized to RPL10 ($n = 3$). Inset is a polymacrylamide gel of qRT-PCR products from the various cells. (*c*) Schematic depicting the sasRNA target loci in the Periostin promoter. (*d*) Periostin expression in sasRNA transfected PC3 cells 72 h post-transfection. The average of triplicate-treated cultures are shown with the standard errors of the mean and *p*-values from a paired *t*-test. Calculations are relative to parent U6M2 plasmid. (*e*) The sequences for the target locus and as6 transcript. $^*p < 0.05$ by two-tailed *t*-test. Error bars indicate s.e.m.

There is a significant correlation between high Periostin expression and poor prognosis in cancer patients [8–12], indicating its therapeutic potential as a gene-silencing target in an effort to reduce metastatic tumour formation.

One pathway for inducing stable long-term transcriptional gene silencing (TGS) involves targeting a genes promoter with small antisense non-coding RNAs (sasRNAs; reviewed in [13]) (figure 1*a*). This process involves the sasRNA recruitment of epigenetic modifying complexes to the RNA targeted locus, ultimately rendering the targeted promoter loci inaccessible to RNA polymerase II (RNAPII) [14–16]. One advantage to sasRNA-directed TGS is that it can produce lasting, stable epigenetic modifications that are inherited by daughter cells [17]. As such we hypothesized that sasRNAs targeted to direct TGS of Periostin could inhibit invasive and metastatic potential of cancer cells. We report here on a sasRNA that can direct TGS of Periostin resulting in the inhibition of cell metastasis. The sasRNA reported here may prove therapeutically relevant as a means for stably silencing Periostin activity and metastasis of tumour cells.

2. Periostin promoter targeting with small antisense non-coding RNAs

Altered expression of Periostin is reported in a number of different forms of cancer [5,7,18]. Endogenous expression of Periostin was tested in four different cell lines. Results from qRT-PCR indicated that PC3 cells from a prostate cancer cell line had the highest expression of Periostin mRNA when compared with the other cell lines assessed (figure 1*b*). Expression was approximately 10 times greater in PC3 cells than that observed in HEK293 cells, which demonstrated the next highest expression. By comparison, CFPAC and HeLa cells showed negligible expression of Periostin. Next, we sought to determine the susceptibility of Periostin to promoter targeted sasRNAs. To determine the ability to direct TGS of Periostin, five sasRNA sequences directed to the Periostin promoter were designed (figure 1*c*), synthesized [19] and

screened in PC3 cells. All five sasRNAs were capable of suppressing Periostin, with one candidate, as6, demonstrating a robust and significant repression of Periostin expression (figure 1*d*). Collectively, these data suggest that Periostin is susceptible to sasRNA-directed TGS, similar to previous observations with other genes [19–31].

3. Periostin knockdown is a result of small antisense non-coding RNA-directed transcriptional gene silencing

The target sequence for as6 in the Periostin promoter overlaps the 5′UTR of Periostin (figure 1*e*) and the observed suppression of Periostin by as6 may be post-transcriptional in nature. In order to confirm that the observed reduction of Periostin mRNA is a result of TGS directed by the sasRNA as6 guide, as has been observed previously with other sasRNA-targeted loci [13,19,32], an in-depth investigation into the nuclear expression of Periostin was conducted. Nuclear run-on analysis of as6-treated cultures indicated that the observed suppression was transcriptional in nature (figure 2*a*), suggesting that as6 targeting of the Periostin 5′UTR is sufficient for modulating the transcriptional activity of Periostin in PC3 cells. SasRNA-directed TGS has been observed to result in transcriptional and epigenetic changes at the sasRNA target locus, that ultimately result in a loss of active forms of RNA polymerase II (RNAPII) and a gain of DNA methyltransferase 3a (DNMT3a) at the targeted promoter (reviewed in [13,14]). To determine if as6 treatment affects the localization of RNAPII or DNMT3a to the Periostin promoter, a chromatin-immunoprecipitation (ChIP) was performed in as6-treated PC3 cells. ChIP pull-down for active forms of RNAPII revealed a significant reduction in RNAPII and approximately fourfold increased enrichment for DNMT3a at the Periostin promoter (figure 2*b*).

Previous studies have observed that TGS can be inhibited by trichostatin A (TSA) and 5′-azacytidine (5′Aza-C), which inhibit the TGS-associated proteins histone deacetylase (HDAC) and DNMT3a, respectively [29]. Treatment of as6-transfected PC3 cells with TSA and 5′Aza-C partially restored Periostin gene expression (figure 2*c*). SasRNA-directed TGS has been observed to also require transcription at the targeted promoter and a promoter-associated RNA [16,33]. When cells were treated with the drug α-amanitin [32,34], an inhibitor of RNAPII, a complete reversion of mRNA expression is observed (figure 2*d*). These data suggest that TGS of Periostin requires RNAPII mediated transcription of the sasRNA guide. Collectively, the observations reported here indicate that as6 is functioning to direct TGS to the Periostin promoter in a manner similar to what has been observed previously with other sasRNA-targeted genes [20,21,32,35] (reviewed in [13,14]).

3.1. as6-mediated phenotypes

In an effort to determine if the sasRNA as6 functionally modulates the metastatic ability of tumour cells, we set out to investigate the observable phenotypic effects of as6 when introduced to tumour cells *in vitro*. During the generation of stable HEK293 and PC3 cell lines expressing the as6 sasRNA using Geneticin (G418) selection, it was discovered over a 6-week period that those cells that survived the selection process in both cell lines were unable to proliferate. As such, the generation of stable cell lines was not feasible, possibly due to the suppression of Periostin by as6. To explore this notion further cell counts in as6-treated versus -untreated PC3 cells were followed up over a 72 h period. Cultures treated with as6 demonstrated a significant reduction in exponential growth (figure 3*a*). Rather than a continued rate of exponential expansion, cell growth was observed to level off and undergo a more linear growth rate after only 3 days of as6 targeting. Notably, by the third day, there were 30% more cells in control samples than in those treated with as6, suggesting as6-directed suppression of Periostin impairs cell growth.

Periostin is known to be involved in cell metastasis [36,37]. To explore the ability of as6 to functionally modulate cell motility and metastasis, a scratch assay was performed on as6- versus control-treated PC3 cells (figure 3*b–d*). Examination of PC3 cell growth into a scratch over a 72-h period demonstrated severely inhibited growth of cells into the scratch in as6-treated cells relative to control cells (figure 3*b*,*d*). After 3 days, control cells were seen to move and proliferate into the scratch, covering over the gap left by the plastic insert (figure 3*b*). By contrast, the as6-treated cells showed very little movement into the provided space (figure 3*d*). The addition of mitomycin C to as6-treated cells caused the scratch to remain completely intact, with no edge distortion occurring, suggesting that the regrowth seen without mitomycin C in as6-treated cells was presumably due to cell migration rather than proliferation (figure 3*c*). These data suggest that as6 sasRNA-directed TGS of Periostin appears to be a potent suppressor of cell migration and is equivalent to the suppression of cell division by mitomycin C.

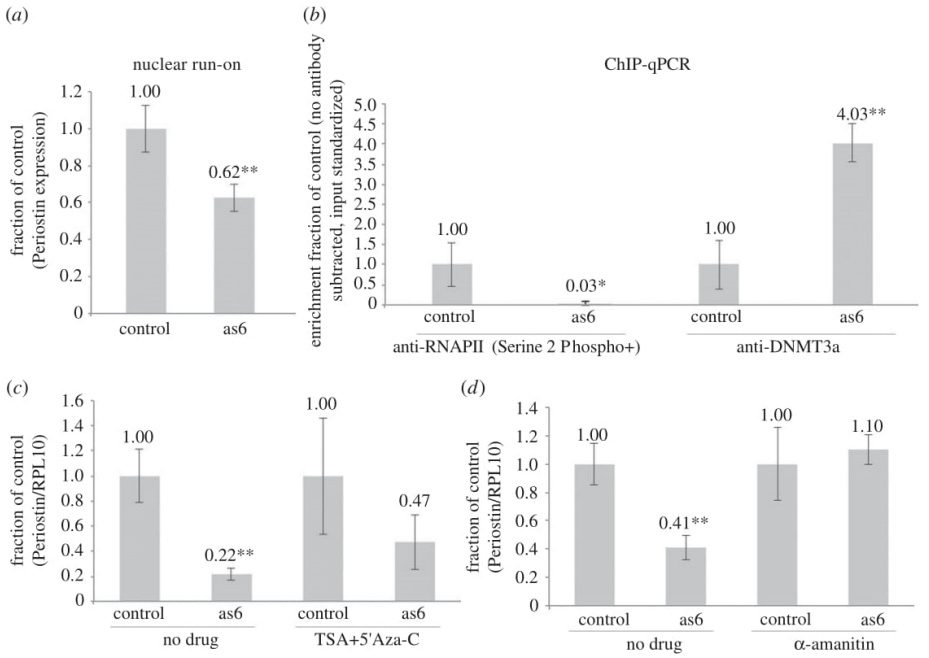


Figure 2. Periostin TGS in the nucleus. (*a*) Nuclear run-on analysis of Periostin transcripts 72 h after transfection with as6, normalized to RPL10 ($n = 3$). Samples are Periostin transcripts 72 h after transfection with as6 ($n = 3$). (*b*) ChIP of RNAPII and DNMT3a at the Periostin promoter was determined 72 h post-transfection. The relative enrichment of RNAPII and DNMT3a was determined following subtraction of no antibody beads alone control and standardized to input. (*c*) Drug treatment using both TSA and Aza-C. Drugs were added every 24 h to cultures treated as described above ($n = 3$), normalized to RPL10. (*d*) Drug treatment using α-amanitin. Drug was added once as described above ($n = 3$). All experiments were performed in PC3 cells. Throughout the figure, $^{*}p < 0.05$, $^{**}p < 0.01$ by two-tailed *t*-test. Calculations are relative to parent U6M2 plasmid. Error bars indicate s.e.m.

Recent studies have presented the concept of endogenous regulatory mechanisms directed by long non-coding RNAs (lncRNAs) in human cells which are responsible for altering the epigenetic state of a target locus, thus effecting the transcriptional expression of a protein coding gene [14,17,38]. This mechanism can be usurped when a sasRNA guide is introduced into the cell, bypassing endogenous lncRNA regulation of the target locus and inducing TGS (figure 1*a*) [13,19,32]. RNA-directed TGS has been observed in human cells before and functions by the sasRNA binding to target low-copy promoter-associated transcripts upstream or overlapping the sasRNA-targeted promoter [16,33], resulting ultimately in the recruitment of epigenetic remodelling complexes to the selected site of interest [13,16] and gene silencing. The advantage of sasRNA-directed TGS over other RNA-based silencing mechanisms, such as RNAi, is that TGS can be long-lasting and heritable. Only a short duration of sasRNA targeting to a gene promoter is required and once DNA methylation is recruited the targeted gene can stay repressed indefinitely [21]. The data presented here suggest that as6 sasRNA is capable of inducing TGS at the 5′UTR/promoter of Periostin and that this targeting has distinct phenotypic effects on cell motility. Collectively, the observations presented here suggest that as6 has the potential to function as an inhibitor of cell motility and possibly as an anti-metastatic molecule that could be highly useful in cancer treatments, although the effect of as6 on cell metastasis remains to be determined and will require further *in vivo* studies.

Previous observations demonstrated a correlation between high expression of Periostin and poor cancer-patient prognosis [8–12]. This is presumably due to Periostin's ability to regulate tumour cell invasion and metastasis, as tumour metastasis is associated with the highest rate of mortality for cancer patients [36,37]. Suppression of Periostin gene expression *in vitro* via indirect methods [39–41], and others directly through RNAi [42–44], demonstrated a direct correlation between Periostin expression and the

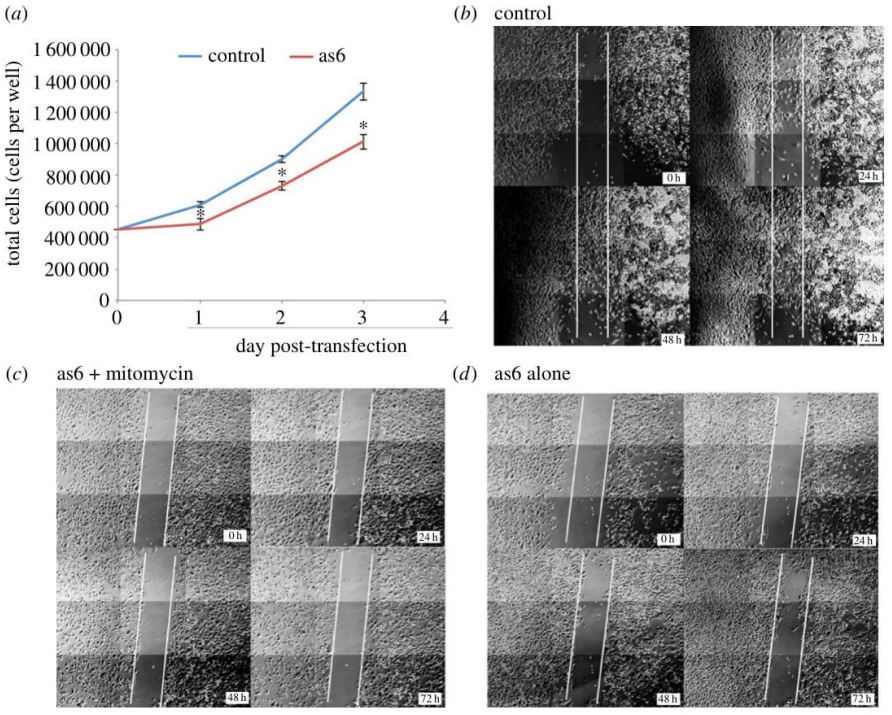


Figure 3. Phenotypic effect of Periostin TGS. (*a*) PC3 cell numbers following as6 treatment. Cells were counted every 24 h post-transfection with as6 or parent U6M2 plasmid. $^{*}p < 0.05$ by two-tailed *t*-test. Error bars indicate s.e.m. (*b–d*) Scratch assay 72 h after transfection with (*b*) parent U6M2 plasmid alone, (*c*) as6 expressing U6M2 treated with mitomycin C and (*d*) as6 expressing U6M2 plasmid alone. Cells were photographed at times 0, 24, 48 and 72 h post-transfection. The white lines indicate the initial scratch.

ability of cells to undergo migration and cell division. While previous studies using RNAi to suppress Periostin clearly demonstrated a loss of cell metastasis, the observed effects are only transient as the targeting of Periostin was at the post-transcriptional level [42–44]. The data presented here suggests that Periostin is susceptible to sasRNA-directed TGS and that it may be possible to stably suppress Periostin [13,17].

Secondary tumours, those that metastasize to new locations, are incapable of surviving without the presence of Periostin [45]. This matches with phenotypic data from the scratch assay and cell count, as well as the inability to generate stable cell lines, as as6 expressing cells would not proliferate and showed very limited ability to migrate, yet they did not undergo apoptosis. Periostin is also required to initiate alteration of the extracellular matrix, creating a niche environment that allows tumour invasion and proliferation [45,46]. The observed inability of the as6-treated cells to divide and migrate suggests the involvement of Periostin in rearranging the extracellular matrix and providing an environment conducive to metastasis. When suppression of this Periostin is induced by TGS, there is no alteration of the surrounding environment to favour cell invasion.

While the TGS induced by as6 was shown to be an effective inhibitor of metastatic and invasive potential of tumour cells, an effective method for introduction of the sasRNA into human patients and specifically to the tumours is required before it can be used to treat cancer patients. New delivery mechanisms that are targeted specifically to diseased cells are required if future treatment employing sasRNA-directed TGS is to become a viable option. However, once the sasRNAs reach their targets they have the distinct potential to provide stable epigenetic modifications to the gene promoter of interest and long-term heritable epigenetic silencing.

Data accessibility. Data are available from the Dryad respository via http://dx.doi.org/10.5061/dryad.dc3m7.

Authors' contributions. N.L. performed all the experiments and wrote the paper, M.C. performed some experiments and K.V.M. designed the experiments and wrote the paper.

Competing interests. The authors declare no competing interests.

Funding. The project was supported by NIH NIAID AI099783-01, NCI CA151574-01 and CA153124-01, and Australian Research Council FT1300100572 to K.V.M.

References

1. Morra L, Moch H. 2011 *Periostin expression and epithelial-mesenchymal transition in cancer: a review and an update*. *Virchows Arch.* **459**, 465–475. (doi:10.1007/s00428-011-1151-5)
2. Morra L *et al.* 2011 Relevance of periostin splice variants in renal cell carcinoma. *Am. J. Pathol.* **179**, 1513–1521. (doi:10.1016/j.ajpath.2011.05.035)
3. Takeshita S, Kikuno R, Tezuka K, Amann E. 1993 Osteoblast-specific factor 2: cloning of a putative bone adhesion protein with homology with the insect protein fasciclin I. *Biochem. J.* **294**, 271–278.
4. Kanno A *et al.* 2008 Periostin, secreted from stromal cells, has biphasic effect on cell migration and correlates with the epithelial to mesenchymal transition of human pancreatic cancer cells. *Int. J. Cancer.* **122**, 2707–2718. (doi:10.1002/ijc.23332)
5. Tischler V *et al.* 2010 Periostin is up-regulated in high grade and high stage prostate cancer. *BMC Cancer.* **10**, 273. (doi:10.1186/1471-2407-10-273)
6. Gritsenko PG, Ilina O, Friedl P. 2012 Interstitial guidance of cancer invasion. *J. Pathol.* **226**, 185–199. (doi:10.1002/path.3031)
7. Kudo Y, Siriwardena BS, Hatano H, Ogawa I, Takata T. 2007 Periostin: novel diagnostic and therapeutic target for cancer. *Histol. Histopathol.* **22**, 1167–1174.
8. Ben QW, Zhao Z, Ge S-F, Zhou J, Yuan F, Yuan Y-Z. 2009 Circulating levels of periostin may help identify patients with more aggressive colorectal cancer. *Int. J. Oncol.* **34**, 821–828. (doi:10.3892/ijo_00000208)
9. Choi KU *et al.* 2011 Lysophosphatidic acid-induced expression of periostin in stromal cells: Prognostic relevance of periostin expression in epithelial ovarian cancer. *Int. J. Cancer* **128**, 332–342. (doi:10.1002/ijc.25341)
10. Forsti A *et al.* 2007 Polymorphisms in the KDR and POSTN genes: association with breast cancer susceptibility and prognosis. *Breast Cancer Res. Treat.* **101**, 83–93. (doi:10.1007/s10549-006-9265-1)
11. Nuzzo PV, Rubagotti A, Zinoli L, Ricci F, Salvi S, Boccardo S, Boccardo F. 2012 Prognostic value of stromal and epithelial periostin expression in human prostate cancer: correlation with clinical pathological features and the risk of biochemical relapse or death. *BMC Cancer* **12**, 625. (doi:10.1186/1471-2407-12-625)
12. Tsunoda T *et al.* 2009 The increased expression of periostin during early stages of prostate cancer and advanced stages of cancer stroma. *Prostate* **69**, 1398–1403. (doi:10.1002/pros.20988)
13. Morris KV. 2009 RNA-directed transcriptional gene silencing and activation in human cells. *Oligonucleotides* **19**, 299–306. (doi:10.1089/oli.2009.0212)
14. Turner AM, Morris KV. 2010 Controlling transcription with noncoding RNAs in mammalian cells. *Biotechniques* **48**, ix–xvi. (doi:10.2144/000113442)
15. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H. 2008 Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451**, 202–206. (doi:10.1038/nature06468)
16. Han J, Kim D, Morris KV. 2007 Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc. Natl Acad. Sci. USA* **104**, 12 422–12 427. (doi:10.1073/pnas.0701635104)
17. Morris KV. 2009 Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells. *Epigenetics* **4**, 296–301. (doi:10.4161/epi.4.5.9282)
18. Ruan K, Bao S, Ouyang G. 2009 The multifaceted role of periostin in tumorigenesis. *Cell Mol. Life Sci.* **66**, 2219–2230. (doi:10.1007/s00018-009-0013-7)
19. Ackley A, Lenox A, Stapleton K, Knowling S, Lu T, Sabir KS, Vogt PK, Morris KV. 2013 An algorithm for generating small RNAs capable of epigenetically modulating transcriptional gene silencing and activation in human cells. *Mol. Ther. Nucleic Acids* **2**, e104. (doi:10.1038/mtna.2013.33)
20. Turner AM, Ackley AM, Matrone MA, Morris KV. 2012 Characterization of an HIV-targeted transcriptional gene-silencing RNA in primary cells. *Hum. Gene Ther.* **23**, 473–483. (doi:10.1089/hum.2011.165)
21. Hawkins PG, Santoso S, Adams C, Anest V, Morris KV. 2009 Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. *Nucleic Acids Res.* **37**, 2984–2995. (doi:10.1093/nar/gkp127)
22. Kim DH, Saetrom P, Snove O, Rossi JJ. 2008 MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl Acad. Sci. USA* **105**, 16 230–16 235. (doi:10.1073/pnas.0808830105)
23. Kim JW, Zhang Y-H, Zern MA, Rossi JJ, Wu J. 2007 Short hairpin RNA causes the methylation of transforming growth factor-beta receptor II promoter and silencing of the target gene in rat hepatic stellate cells. *Biochem. Biophys. Res. Commun.* **359**, 292–297. (doi:10.1016/j.bbrc.2007.05.080)
24. Lim HG, Suzuki K, Cooper DA, Kelleher AD. 2008 Promoter-targeted siRNAs induce gene silencing of simian immunodeficiency virus (SIV) infection in vitro. *Mol. Ther.* **16**, 565–570. (doi:10.1038/sj.mt.6300380)
25. Roberts TC, Andaloussi SEL, Morris KV, McClorey G, Wood MJA. 2012 Small RNA-mediated epigenetic myostatin silencing. *Mol. Ther. Nucleic Acids.* **1**, e23. (doi:10.1038/mtna.2012.15)
26. Suzuki K *et al.* 2011 Transcriptional gene silencing of HIV-1 through promoter targeted RNA is highly specific. *RNA Biol.* **8**, 1035–1046. (doi:10.4161/rna.8.6.16264)
27. Suzuki K, Juelich T, Lim H, Ishida T, Watanebe T, Cooper DA, Rao S, Kelleher AD. 2008 Closed chromatin architecture is induced by an RNA duplex targeting the HIV-1 promoter region. *J. Biol. Chem.* **283**, 23 353–23 363. (doi:10.1074/jbc.M709651200)
28. Singh A, Palanichamy JK, Ramalingam P, Kassab MA, Bhagat M, Andrabi R, Luthra K, Sinha S, Chattopadhyay P. 2013 Long-term suppression of HIV-1C virus production in human peripheral blood mononuclear cells by LTR heterochromatization with a short double-stranded RNA. *J. Antimicrob. Chemother.* **69**, 404–415. (doi:10.1093/jac/dkt348)
29. Morris KV. 2004 Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289–1292. (doi:10.1126/science.1101372)
30. Janowski BA, Huffman KE, Schwartz JC, Ram R, Hardy D, Shames DS, Minna JD, Corey DR. 2005 Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat. Chem. Biol.* **1**, 210–215. (doi:10.1038/nchembio724)
31. Janowski BA, Huffman KE, Schwartz JC, Ram R, Nordsell R, Shames DS, Minna JD, Corey DR. 2006 Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat. Struct. Mol. Biol.* **13**, 787–792. (doi:10.1038/nsmb1140)
32. Weinberg MS. 2006 The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA* **12**, 256–262. (doi:10.1261/rna.2235106)
33. Napoli S, Pastori C, Magistri M, Carbone GM, Catapano CV. 2009 Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J.* **28**, 1708–1719. (doi:10.1038/emboj.2009.139)
34. Kato H. 2005 RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**, 467–469. (doi:10.1126/science.1114955)
35. Turner AM, De La Cruz J, Morris KV. 2009 Mobilization-competent lentiviral vector-mediated sustained transcriptional modulation of HIV-1 expression. *Mol. Ther.* **17**, 360–368. (doi:10.1038/mt.2008.268)
36. Wang Z, Ouyang G. 2012 Periostin: a bridge between cancer stem cells and their metastatic niche. *Cell Stem Cell* **10**, 111–112. (doi:10.1016/j.stem.2012.01.002)
37. Ben QW, Jin X-L, Liu J, Cai X, Yuan F, Yuan Y-Z. 2011 Periostin, a matrix specific protein, is associated with proliferation and invasion of pancreatic cancer. *Oncol. Rep.* **25**, 709–716. (doi:10.3892/or.2011.1140)
38. Morris KV, Mattick JS. 2014 The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–437. (doi:10.1038/nrg3722)
39. Deraz EM *et al.* 2011 MMP-10/stromelysin-2 promotes invasion of head and neck cancer. *PLoS ONE* **6**, e25438. (doi:10.1371/journal.pone.0025438)
40. Kudo Y. 2006 Periostin promotes invasion and anchorage-independent growth in the metastatic process of head and neck cancer. *Cancer Res.* **66**, 6928–6935. (doi:10.1158/0008-5472.CAN-05-4540)

41. Malanchi I, Huelsken J. 2009 Cancer stem cells: never Wnt away from the niche. *Curr. Opin. Oncol.* **21**, 41–46. (doi:10.1097/CCO.0b013e32831d1faf)
42. Sun C, Zhao X, Xu K, Gong J, Liu W, Ding W, Gou Y, Xia G, Ding Q. 2011 Periostin: a promising target of therapeutical intervention for prostate cancer. *J. Transl. Med.* **9**, 99. (doi:10.1186/1479-5876-9-99)
43. Liu Y, Liu BA. 2011 Enhanced proliferation, invasion, and epithelial-mesenchymal transition of nicotine-promoted gastric cancer by periostin. *World J Gastroenterol.* **17**, 2674–2680. (doi:10.3748/wjg.v17.i21.2674)
44. Xu D, Xu H, Ren Y, Liu C, Wang X, Zhang H, Lu P. 2012 Cancer stem cell-related gene periostin: a novel prognostic marker for breast cancer. *PLoS ONE* **7**, e46670. (doi:10.1371/journal.pone.0046670)
45. Malanchi I, Santamaria-Martínez 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**, 85–89. (doi:10.1038/nature10694)
46. Kim CJ, Yoshioka N, Tambe Y, Kushima R, Okada Y, Inoue H. 2005 Periostin is down-regulated in high grade human bladder cancers and suppresses *in vitro* cell invasiveness and *in vivo* metastasis of cancer cells. *Int. J. Cancer* **117**, 51–58. (doi:10.1002/ijc.21120)

7.2 Paper II

PNAS

The molecular dynamics of long noncoding RNA control of transcription in PTEN and its pseudogene

Nicholas Lister[a,1], Galina Shevchenko[b,1], James L. Walshe[c], Jessica Groen[a], Per Johnsson[d], Linda Vidarsdóttir[e], Dan Grander[e], Sandro F. Ataide[c], and Kevin V. Morris[b,2]

[a]The University of New South Wales, Biotechnology and Biomedical Sciences, Sydney, NSW 2052, Australia; [b]Center for Gene Therapy, City of Hope–Beckman Research Institute at the City of Hope, Duarte, CA 91010; [c]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia; [d]Department of Cell and Molecular Biology, Ludwig Institute for Cancer Research, Karolinska Institutet, 171 65 Stockholm, Sweden; and [e]Department of Oncology-Pathology, Cancer Center Karolinska, R8:03, Karolinska Hospital and Institute, S-171 76 Stockholm, Sweden

Edited by Jeannie T. Lee, Harvard Medical School Massachusetts General Hospital, Boston, MA, and approved August 2, 2017 (received for review January 4, 2017)

RNA has been found to interact with chromatin and modulate gene transcription. In human cells, little is known about how long noncoding RNAs (lncRNAs) interact with target loci in the context of chromatin. We find here, using the phosphatase and tensin homolog (PTEN) pseudogene as a model system, that antisense lncRNAs interact first with a 5′ UTR-containing promoter-spanning transcript, which is then followed by the recruitment of DNA methyltransferase 3a (DNMT3a), ultimately resulting in the transcriptional and epigenetic control of gene expression. Moreover, we find that the lncRNA and promoter-spanning transcript interaction are based on a combination of structural and sequence components of the antisense lncRNA. These observations suggest, on the basis of this one example, that evolutionary pressures may be placed on RNA structure more so than sequence conservation. Collectively, the observations presented here suggest a much more complex and vibrant RNA regulatory world may be operative in the regulation of gene expression.

PTEN | DNMT3a | pseudogene | PTENpg1 | epigenetic

The phosphatase and tensin homolog (PTEN) is a protein encoded on chromosome 10 by the *PTEN* gene and is a negative regulator of the oncogenic PI3K-protein kinase B (Akt) pathway. The *PTEN* gene is mutated and epigenetically inactivated in a diverse range of cancers (1). This gene is of particular interest, as emerging studies have shown that a pseudogene, PTENpg1, is actively involved in the regulation of *PTEN*. These observations suggest an underappreciation of the complexity involved in gene regulation. To date, thousands of pseudogenes have been identified in humans, including many disease-associated genes such as TP53 (2), BRCA1 (3), OCT4 (4–6), and PTEN (7). PTEN has a single pseudogene in the human genome, PTENpg1 (also called PTENp1, PTENΨ), which is encoded by chromosome 9 (8). PTENpg1 posttranscriptionally regulates *PTEN* expression by acting as a miRNA sponge to PTEN-targeting miRNAs (9). Recent studies have indicated the presence of antisense RNAs (asRNAs) derived from the PTENpg1 promoter locus (10). Several different isoforms of this antisense, named α and β, have been identified with transcription arising from the bidirectional PTENpg1 promoter, and one variant, PTENpg1 asRNA α, is found to modulate PTEN transcription via the recruitment of chromatin-modifying complexes EZH2 and DNMT3a (10). These proteins are actively recruited to the promoter by PTENpg1 asRNA α and cause chromatin condensation, and subsequently a reduction in *PTEN* expression (10).

It is noteworthy that *PTEN* and PTENpg1 are localized on different chromosomes, and the putative in trans acting mechanism by which the PTENpg1 asRNA α interacts with the *PTEN* promoter has not been determined. We find here that PTENpg1 asRNA α targeting DNMT3a to the *PTEN* promoter requires transcription of PTEN, specifically at the 5′UTR region containing homology to the PTENpg1 asRNA α transcript, and that this RNA can target the PTEN promoter in the absence of DNMT3a.

Results

Detection and Function of 5′ UTR PTEN Promoter Transcripts. Previous studies with small noncoding RNAs (ncRNAs) demonstrated that an expressed low-copy transcript spanning the 5′ UTR of protein coding genes, designated a promoter-associated RNA (paRNA), was required for small ncRNA-directed epigenetic regulation in human cells (11–13). Transcriptomic data suggest there are several expressed sequence tags (ESTs) spanning the *PTEN* promoter [including BG772190 (Fig. 1*A* and Table S1) and DA005202, DA676942, and CN413383]. To detect the presence of these transcripts at the PTEN promoter, and to examine to what extent that they may play a role in PTENpg1 asRNA α regulation of *PTEN* transcription, strand-specific directional RT-PCR was performed on HeLa total RNA, using the PTENproR1 reverse primer only (Fig. S1*A*). This assay allows for any RNA transcripts spanning the *PTEN* promoter in a sense orientation to be discerned. Notably, a product was observed spanning this region of *PTEN* (Fig. 1*B*). Unfortunately, it is virtually impossible to disentangle to what extent this *PTEN* promoter-spanning transcript (Fig. 1*B*) is either a unique low-copy transcript or the 5′ UTR of *PTEN* mRNAs (Fig. 1*A*). Next, to interrogate to what extent the PTEN 5′ UTR-containing transcripts (paRNA/5′UTR) are involved in PTENpg1 asRNA α regulation of *PTEN*, suppression of this transcript was carried out using single-stranded antisense phosphorothioate oligodeoxynucleotides

Significance

In recent years, noncoding RNA transcripts have been found to interact with genes and modulate their ability to be transcribed and made into protein. Here we uncover many of the mechanistic underpinnings involved in how noncoding RNAs control gene transcription. Notably, we find that noncoding RNA control of gene transcription is based on a combination of structural and sequence components of the noncoding RNA and targeted gene. Collectively, the observations presented here suggest that a much more complex and vibrant RNA regulatory world is operative in gene expression and evolution of the genome.

Author contributions: G.S., J.L.W., P.J., D.G., S.F.A., and K.V.M. designed research; N.L., G.S., J.L.W., J.G., P.J., and L.V. performed research; D.G. contributed new reagents/analytic tools; N.L., G.S., L.V., S.F.A., and K.V.M. analyzed data; and N.L., G.S., S.F.A., and K.V.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]N.L. and G.S. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: kmorris@coh.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1621490114/-/DCSupplemental.

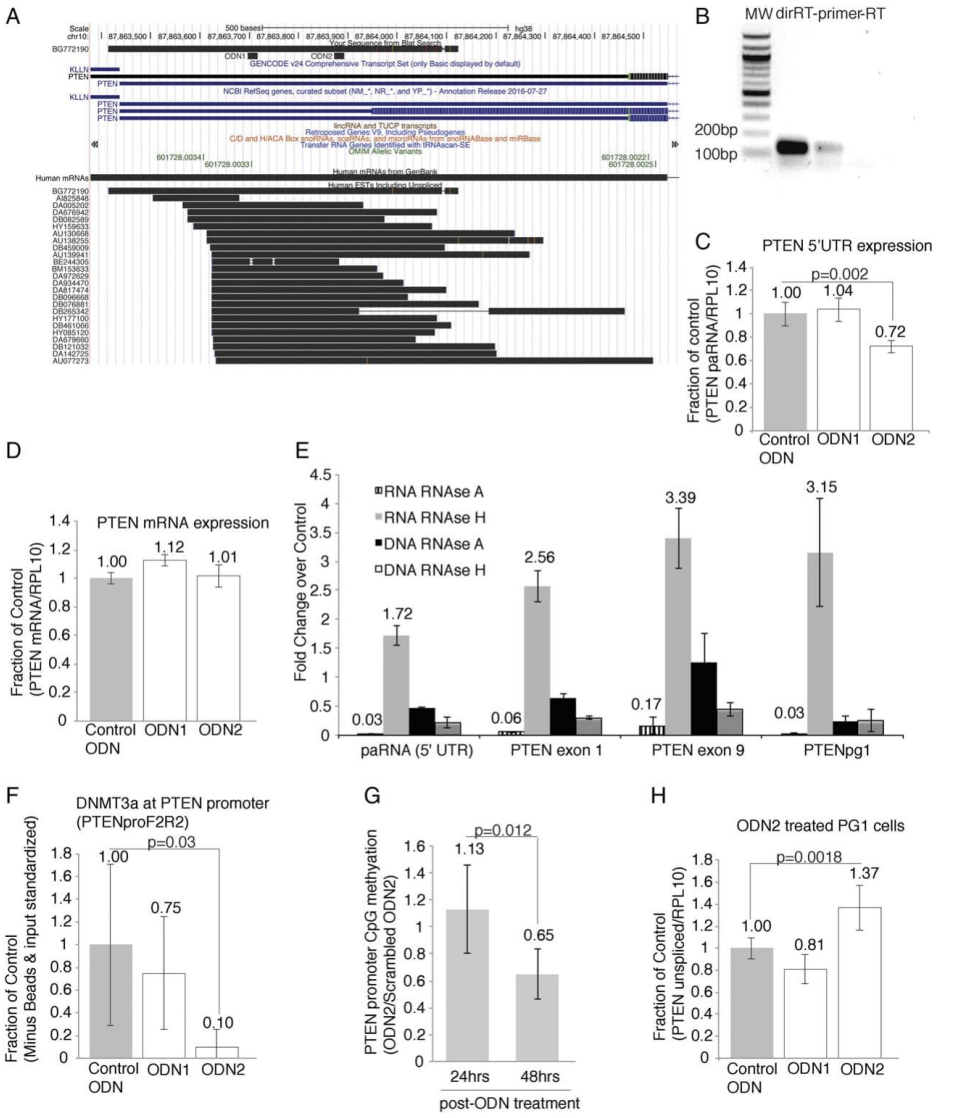


Fig. 1. The PTEN promoter transcriptional landscape. (*A*) A screen shot from the University of California, Santa Cruz (UCSC) genome browser depicting several ESTs at the PTEN promoter/5′ UTR. EST BG772190 is also shown, which directly overlaps the PTENpg1 asRNA exon 1 target site (10). (*B*) Detection of a paRNA (e.g., EST BG772190) by directional RT-PCR were run on a 2% agarose gel. (*Left*) A 100-bp ladder. PCR products are HeLa directional RT (diRT), HeLa no primer (noPrimer), and No template RT (noRT). (*C* and *D*) The effects of antisense phosphorothioate oligodeoxynucleotide targeting of the PTEN pRNA (EST BG772190). (*C*) Expression of the PTEN 5′UTR (e.g., paRNA/EST BG772190) and (*D*) PTEN mRNA 48 h posttransfection with ODNs. (*E*) Biotin ODN2 binds predominantly PTEN 5′UTR-containing transcripts and not the PTEN gene. Biotin ODN2 IP was performed on 293HEK cells, and RNase A vs. RNase H sensitivity was determined. (*F*) Effects of ODN treatment on DNMT3a localization to the PTEN promoter, as determined by ChIP 48 h posttransfection with ODNs. (*G*) PTEN promoter CpG methylation, as determined by methyl-cytosine-specific restriction enzyme McrBc treatment and qPCR in ODN2 and ODN2 Scrambled Control-treated 293HEK cells 24 and 48 h posttransfection. The averages of triplicate-treated cultures are shown, and the *P* value from a paired *t* test, a single representative experiment, is shown. (*H*) The effects of ODN treatment on expression of unspliced variants of PTEN in stable 293HEK cells that overexpress PTENpg1 asRNA exon1 (PG1 cells). PTEN RNA expression was determined by qRT-PCR with PTEN F3/R3 primers, which specifically detect unspliced forms of PTEN. For *C*, *D*, *F*, and *H*, triplicate-treated samples are shown as a fraction of the control ODN ± SE of mean (SEM) after normalization to RPL10, and for *E*, duplicates with the ranges shown from a single representative experiment. *P* values from a paired two-tailed *t* test are also shown.

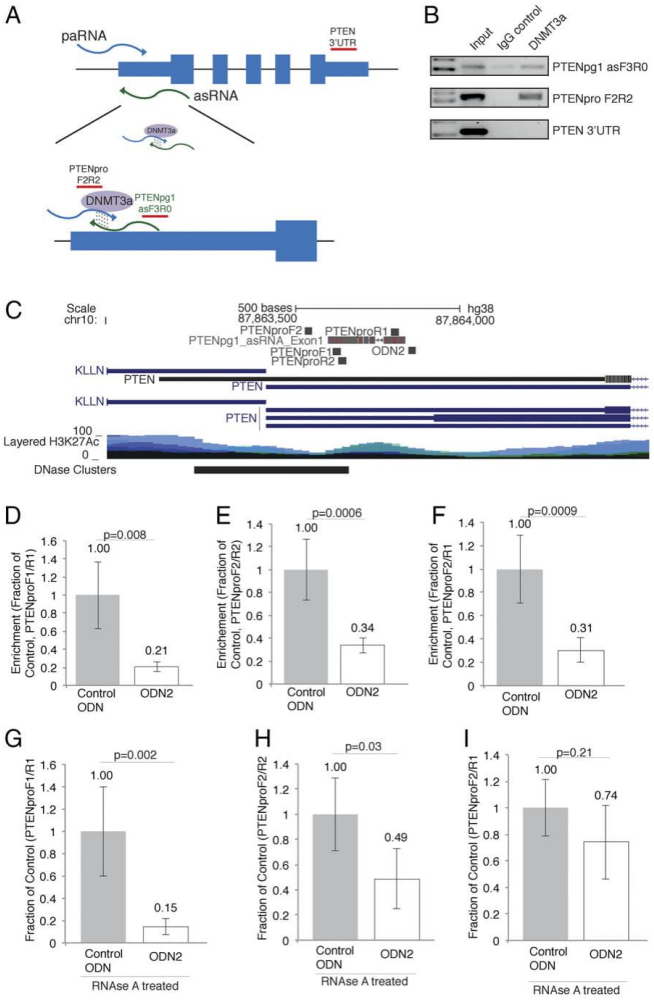


Fig. 2. The effects of ODN2 treatment on PTENpg1 asRNA exon1 binding to and regulation of PTEN. (*A*) A schematic is shown depicting the predicted PTEN paRNA and PTENpg1 asRNA exon 1 interacting region relative to PTEN, along with the RNA:RNA interacting domain and presumed DNMT3a interaction (10). (*B*) ChIP was carried out for DNMT3a in 293HEK cells, and those transcripts associated with DNMT3a were determined by semiqRT-PCR. DNMT3a was observed to interact with both PTENpg1 asRNA (asF3R0) and paRNA/EST BG772190 (PTENpro F2R2) loci, but not the 3′ UTR of PTEN. The PTENpro F2/R2 primer set is specific for PTEN-5′UTR paRNAs (e.g., EST BG772190), whereas PTENpg1 asF3R0 is specific for PTEN pg1-encoded transcripts. A primer set targeting the PTEN 3′UTR is provided as a control. (*C*) A UCSC genome browser screenshot of the PTENpg1 antisense RNA exon 1 target site in the PTEN promoter is shown along with various primers used to interrogate PTENpg1 asRNA exon 1 interactions with the PTEN promoter. (*D–I*) RIP analysis of biotin PTENpg1 asRNA α interactions with the PTEN promoter in the presence of ODN2. (*D–F*) ODN2 treatment blocks biotin dUTP-containing PTENpg1 asRNA exon1 binding to the PTEN promoter. Biotin-labeled PTENpg1 asRNA exon 1 was cotransfected with ODN2 into 293HEK cells, and localization to the PTEN promoter determined by ChIP. (*D*) Primer set PTENproF1/R1, (*E*) PTENpro F2/R2, and (*F*) PTENproF2/R1 were used to detect the loss of PTENpg1 asRNA α exon 1 binding after ODN2 treatment to various loci in the PTEN promoter. (*G–I*) Mapping of RNA:RNA interacting loci at the PTEN promoter. ODN2 treatment results in reduced PTENpg1 asRNA exon1 interactions with PTEN paRNA/EST BG772190 at the (*G*) PTENproF1/R1 and (*H*) PTENproF2/R2 loci, but not the (*I*) PTENproF2/R1 locus. For *D–I*, the averages of triplicate-treated cultures are shown with the SE of mean and *P* values from a paired two-tailed *t* test.

(ODNs) (14) (Fig. S1*B*). ODNs allow strand-specific interactions to be targeted and blocked, thereby disrupting only the sense-stranded paRNA/5′UTR for *PTEN* (10–12). We observed here that targeting the paRNA/5′UTR with ODN2 resulted in suppression of the *PTEN* promoter-associated 5′ UTR transcripts (Fig. 1*C*), but had little to no effect on downstream *PTEN* mRNA expression (Fig. 1*D*), suggesting there may be unique 5′ UTR-associated transcripts overlapping the PTEN promoter, similar to previous observations with small RNA-targeted transcriptional regulatory mechanisms (11, 12).

To distinguish whether ODN2 was targeting the PTEN gene or a unique 5′ UTR-associated transcript, we used biotin-labeled ODN2 to immunoprecipitate (IP) ODN2-bound nucleic acids. By using RNase A and RNase H treatment, we determined that ODN2 binds a PTEN 5′ UTR-associated transcript, which interestingly extends the full length of the PTEN mRNA (Fig. 1*E*), similar to previous observations involved in the mechanism of small noncoding RNA transcriptional gene silencing (11, 12). Notably, although ODN2 also appeared to bind to the PTENpg1 transcript (Fig. 1*E*), this interaction did not appear to affect PTENpg1 sense and antisense expression (Fig. S2), suggesting ODN2 interacts with a full-length 5′ UTR-containing PTEN transcript, which we refer to as paRNA/5′UTR.

The pathway of PTENpg1 regulation of *PTEN* involves the recruitment of DNMT3a by exon 1 of the PTENpg1 asRNA α transcript to the *PTEN* promoter (10). We find here that ODN2 targeting of the *PTEN* paRNA/5′UTR results in a loss of DNMT3a at the *PTEN* promoter (Fig. 1*F*) and a reduction in CpG methylation at one locus in the PTEN promoter (Fig. 1*G*). This effect of ODN2 on *PTEN* appeared to be the result of ODN2 blocking the activity of the PTENpg1 asRNA α exon 1. Blocking this transcript results in preventing the transcriptional regulation of the *PTEN* promoter by

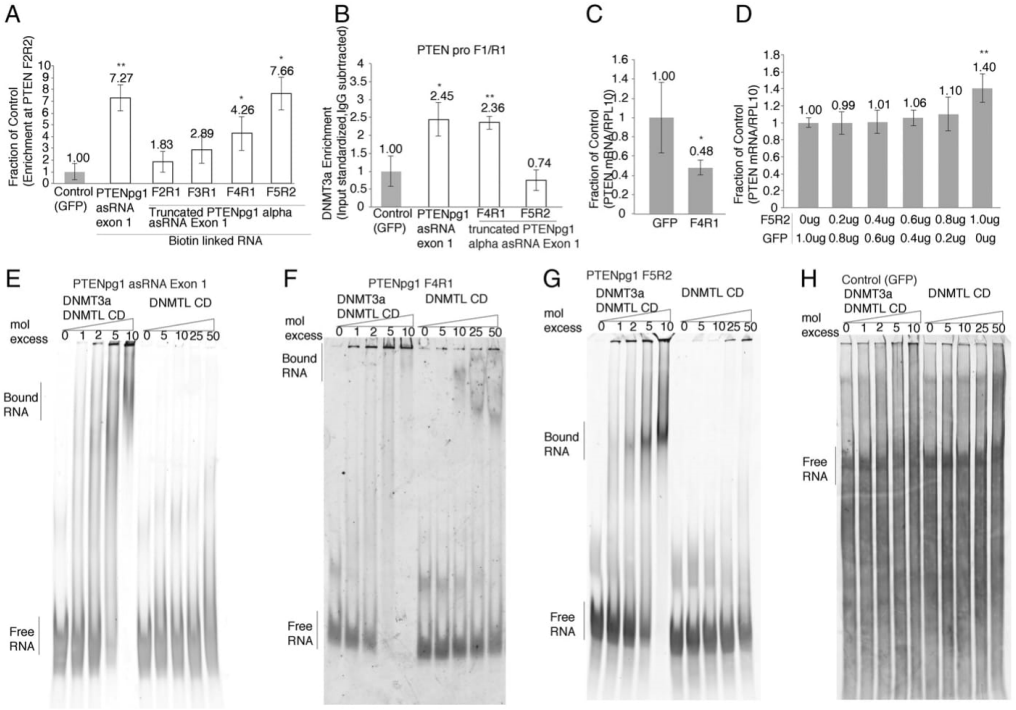


Fig. 3. Characterization of PTENpg1 asRNA a exon 1 and truncated variants involvement in the recruitment of DNMT3a to the PTEN promoter. (*A*) Enrichment of biotin-labeled PTENpg1 asRNA alpha full-length and truncated variants by RIP at the PTEN promoter (F2/R2 locus) after subtraction of beads alone and standardized to inputs. (*B*) Ability of PTENpg1 asRNA exon 1 and truncated variants to direct DNMT3a to the PTEN promoter, as determined by ChIP. (*C*) The PTEN F4R1 variant suppresses PTEN mRNA expression. Either the control pcDNA3.1-GFP or pcDNA-F4R1 were transfected in triplicate into 293HEK cells and PTEN mRNA expression determined 72 h later by qRT-PCR. (*D*) The PTEN F5R2 variant activates PTEN expression based on qRT-PCR analysis of transcript expression. Dose-dependent transfection of F5R2 activation of PTEN mRNA expression was observed in 293HEK cells transfected in triplicate with increasing concentrations of either pcDNA3.1-F5R2 or control pcDNA3.1-GFP. (*E–H*) EMSA analysis of (*E*) PTENpg1 asRNA exon 1, (*F*) truncated F4R1, (*G*) truncated F5R2, and (*H*) Control GFP binding to DNMT3a in the presence of DNMTCD. The reconstituted system Dnmt3a/DnmtL-CD in vitro requires the presence of DNMTL-CD, which alone is unable to bind RNA (even at 50× molecular excess), indicating that the observed RNA binding in the DNMT3a/DNMTL-CD complex comes from DNMT3a. For *A* and *B*, the averages of triplicate transfected 293HEK cells from a single representative experiment are shown with the SEM and *P* values from a two-sided *t* test. $*P < 0.05$; $**P < 0.01$. For *D* (**), a one-way ANOVA was carried out with the f-ratio value of 10.38725; the *P* value is 0.00913.

PTENpg1 asRNA α, as ODN2 treatment in stable PTENpg1 asRNA α exon 1 overexpressing cells demonstrated increased expression of unspliced forms of PTEN (Fig. 1*H*). Collectively, these observations, along with others (10), suggest *PTEN* promoter-associated 5′ UTR-containing transcripts are required for PTENpg1 asRNA α- and DNMT3a-based epigenetic regulation of *PTEN*.

PTEN paRNA/5′UTR Interacts Directly with PTENpg1 asRNA α Exon 1. Previous studies with small antisense RNAs targeted to gene promoters have demonstrated that the small antisense RNAs interact directly with a transcript at the promoter (11, 12). The observations presented here (Fig. 1), along with previous studies (10), suggest the long noncoding RNA (lncRNA) PTENpg1 asRNA α regulates PTEN transcription by interactions with a sense-stranded transcript or elongated 5′UTR (paRNA/5′UTR) that essentially spans the *PTEN* promoter. To interrogate whether the PTEN promoter-associated transcript interacts directly with PTENpg1 asRNA α and DNMT3a, an immunoprecipitation of DNMT3a was carried out, followed by qRT-PCR with primers specific to each transcript (Fig. 2*A*). We find, using this technique, that DNMT3a interacts with both PTENpg1 asRNA alpha exon 1 and a *PTEN* paRNA/5′UTR promoter-associated transcript (Fig. 2 *A* and *B*). To interrogate whether the PTEN promoter-associated transcript interacts directly with PTENpg1 asRNA α, biotin-labeled PTENpg1 asRNA α exon1 was generated and cotransfected with ODN2 into 293HEK cells, and the effects on biotin-labeled PTENpg1 asRNA α exon1 binding to the *PTEN* promoter were determined by RNA immunoprecipitation (RIP; ref. 4 and Fig. 2 *D–I*). ODN2 treatment, relative to controls, abrogated the binding of PTENpg1 asRNA α exon 1 to the PTEN promoter (Fig. 2 *D–F*). To more clearly map the interacting domain of the PTEN 5′ UTR promoter-associated transcript (paRNA/5′UTR) and PTENpg1 asRNA α, RNase A treatment was carried out on the ODN2 and biotin PTENpg1 asRNA α exon 1 cotransfected 293HEK cells. Of those regions bound to the PTEN promoter in the presence of ODN2, one particular region spanning the PTENpg1 asRNA α exon 1 homologous target region and ∼150 nucleotides upstream in the *PTEN* promoter (Fig. 2*G*) appeared to exhibit the most abundant RNA:RNA interactions, based on RNase A sensitivity, relative to the other loci assessed

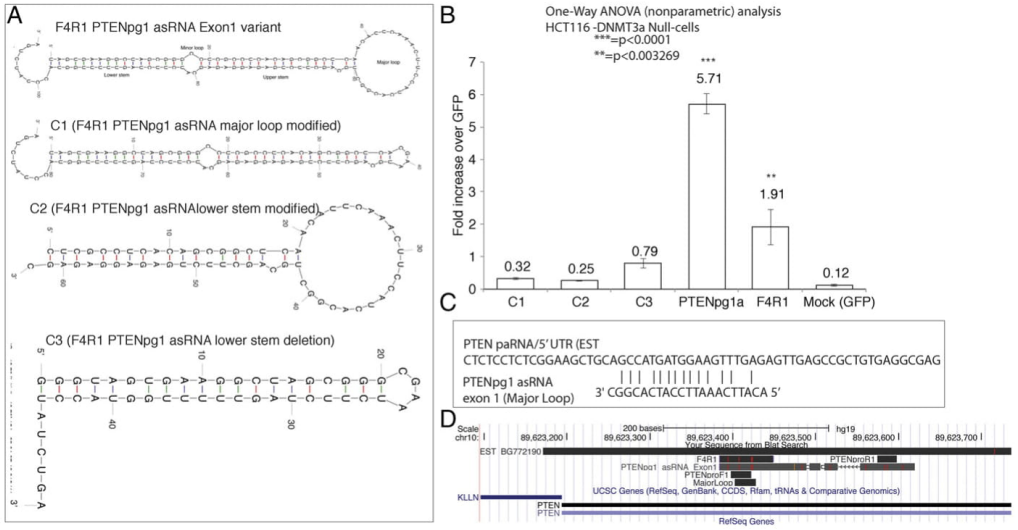


Fig. 4. Characterization of the PTENpg1 asRNA F4R1 variant functional domain. (*A*) Various truncations of F4R1 were developed and cloned into the pcDNA3.1 expression vector and screened for binding to the PTEN promoter. M-Fold analysis of the predicted RNAs is shown. (*B*) Biotin-labeled PTENpg1 asRNA exon1, F4R1, and F4R1 variants C1-F4R1 PTENpg1 asRNA Δ major loop, C2-F4R1 PTENpg1 asRNA Δ lower stem, and C3-F4R1 PTENpg1 as RNA Δ upper stem and major loop binding to the PTEN promoter in the DNMT3a deleted HCT116 cultures. The average of triplicate-treated cultures are shown with *P* values from a one-way ANOVA (nonparametric) analysis. ***$P < 0.0001$; **$P < 0.003269$. (*C*) PTENpg1 asRNA exon 1 and F4R1 major loop interacting locus in the PTEN paRNA (EST BG772190)/5′ UTR of PTEN). (*D*) A screen shot from the UCSC genome browser showing the PTENproF1/R1 region in which PTENpg1 asRNA exon 1 and F4R1 interact with EST BG772190. The major loop interacting region is also shown.

(Fig. 2 *H* and *I*), suggesting this region (Fig. 2*C*; PTENproF1/R1) may be where the strongest localization of those RNA:RNA interactions required for PTENpg1 asRNA α regulation of PTEN transcription occur.

Previous observations have indicated PTENpg1 asRNA α exon 1 binds and directs DNMT3a to the *PTEN* promoter (10). To determine the parameters and particular region in PTENpg1 asRNA α exon 1 involved in binding to the PTEN promoter, various truncations of PTENpg1 asRNA α exon 1 were generated as biotin-labeled transcripts (Fig. S3 and Table S2). These truncated PTENpg1 asRNA α exon 1 variants (Fig. S3*A* and Table S2) were transfected into 293HEK cells, and localization to the *PTEN* promoter determined by chromatin immunoprecipitation (ChIP). Two truncated fragments appeared to localize to the *PTEN* promoter, F4R1 and F5R2, as well as the control full-length PTENpg1 asRNA α exon 1 (Fig. 3*A* and Fig. S3). However, when these deletion constructs were assessed for their ability to direct DNMT3a to the *PTEN* promoter, only the full-length PTENpg1 asRNA α exon 1 and F4R1 variant was functionally capable of directing DNMT3a to the *PTEN* promoter (Fig. 3*B*). Interestingly, the F4R1 truncated variant was able to repress PTEN mRNA expression (Fig. 3*C*), similar to previous observations with PTENpg1 asRNA exon 1 (10), whereas the F5R2 variant resulted in a dose-dependent increase in PTEN expression (Fig. 3*D*). Indeed, both PTENpg1 asRNA α exon 1 and F4R1 were found to interact directly with DNMT3a/DNMTL-CD in vitro (Fig. 3 *E* and *F* and Figs. S4 and S5), relative to the control GFP RNA (Fig. 3*H*), as observed in electrophoretic mobility shift assays (EMSA). Interestingly, the F5R2 variant was also found to bind DNMT3a/DNMTL-CD in vitro (Fig. 3*G*), but, unlike the full-length PTENpg1 asRNA α exon 1 or F4R1 variants, was unable to direct DNMT3a to the PTEN promoter (Fig. 3*B*). This is an interesting observation, as the F5R2 variant appears to bind the PTEN promoter (Fig. 3*A*) or DNMT3a (Fig. 3*G*), but not both (Fig. 3*B*), suggesting, based on this single observation, that some antisense lncRNAs may target genes in the absence of DNMT3a or bind DNMT3a and block endogenous recruitment to their intended target. In the case presented here with the PTEN locus, the dose-dependent overexpression of the F5R2 variant appeared to result in active increases in PTEN expression (Fig. 3*D*).

To interrogate this notion further and determine the requirement of an observed major loop (5′-ACAUUCAAA-CUUCCAUCACGGC-3′) that was found in both the PTEN pg1 asRNA exon 1 (Fig. S3*B*) and F4R1 variant (Fig. 4*A* and Fig. S3*C*), but not in the F5R2 variant (Fig. S3*D*), in binding to the PTEN promoter, we generated several truncations of F4R1 (Fig. 4*A* and Table S3) and determined their respective ability, along with the full-length F4R1 control, to bind the PTEN promoter in the absence of DNMT3a. Interestingly, only the controls full-length F4R1 and PTENpg1 asRNA exon 1 were able to bind the PTEN promoter in the absence of DNMT3a (Fig. 4*B*), suggesting the major loop observed in both the PTENpg1 asRNA exon 1 and the F4R1 variant is required for PTENpg1 asRNA targeting of the PTEN promoter in the absence of DNMT3a (Fig. 4 *C* and *D*).

Discussion

The observations presented here suggest *PTEN* 5′ UTR promoter-associated transcripts are involved in PTENpg1 asRNA exon 1-directed epigenetic regulation of *PTEN*. Previous studies have detected promoter-associated transcripts (11, 15), which are required for small antisense RNAs to guide DNMT3a and direct transcriptional gene silencing in human cells (11–13). These promoter-associated transcripts are thought to be low abundant mRNAs that contain elongated 5′ UTRs that can be detected through

directional RT-PCR (11, 12) and public deep sequencing (Fig. 1*A* and Table S1). The data presented here juxtaposed with previous observations suggest that a mechanism of action is active in human cells whereby RNA:RNA interactions occur at chromatin to facilitate the recruitment of epigenetic regulatory protein complexes (16). Building on observations that PTENpg1 antisense RNA α exon 1 is an active transcriptional and epigenetic modulator of *PTEN* (10), we find here that the localization of PTENpg1 asRNA α exon 1 and a truncated variant F4R1 to the *PTEN* promoter requires a PTEN promoter-associated RNA (paRNA/5′UTR) and involves a conserved major loop domain (5′-ACAUUCAAACUUCCAUCACGGC-3′) to successfully direct DNMT3a to the PTEN promoter (Fig. 4). Interestingly, through deletion studies, one sequence appears to be the main modulator involved in the ability of PTENpg1 asRNA exon1 or F4R1 to target DNMT3a to the PTEN promoter, which was not retained in the defective F5R2 variant. This sequence (Fig. 4*C*) appears in both the PTENpg1 asRNA exon 1 and F4R1 variants and maps directly to a region that was observed previously to exhibit high levels of DNMT3a and PTENpg1 asRNA exon 1 binding (10) (Fig. 4*D*).

The observations presented here suggest that an RNA:RNA interaction is involved in PTENpg1 asRNA exon1 targeting of the *PTEN* promoter and that the conserved domain required for this interaction consists of the major loop domain (5′-ACAUUCAAACUUCCAUCACGGC-3′) interacting with the *PTEN* paRNA/5′UTR (Fig. 4 *C* and *D*). This loop appears to be required for localization of the PTENpg1 asRNA exon 1 and F4R1 variant transcripts to the PTEN promoter, along with the entire stem present in the F4R1. When this loop is altered, as is the case with the F5R2 variant, there appears to be a loss of localization of the transcript to the PTEN promoter, whereas an ability to interact with DNMT3a remained intact. Collectively, the dichotomous observations presented here among PTENpg1 asRNA exon1, F4R1, and the F5R2 variants suggest the major loop domain (Fig. 4 *A* and *C* and Fig. S3 *B* and *C*), in combination with a longer stem found in both PTENpg1 alpha asRNA exon 1 and the F4R1 variant, interacts directly with DNMT3a to direct transcriptional and epigenetic silencing of *PTEN*. Such observations may support the notion that lncRNAs and their putative evolutionary conservation may be more contingent on a combination of both structure and sequence (17). Collectively, the observations presented here expand our understanding of endogenous lncRNA networks in human cells and suggest RNA:RNA interactions, particularly at gene promoters, may be mechanistically relevant in lncRNA regulation of protein-coding gene expression. An understanding of this emerging mode of gene and epigenetic regulation could prove useful in the development of targeted therapeutics to disrupt or augment transcriptional regulatory networks in humans.

Materials and Methods

Biotin ODN2 Immunoprecipitation. As described (12, 18), 5′ biotin-labeled ODN2 or scrambled control (Table S4) was transfected into 293HEK cells. Cells were harvested 48 h posttransfection and permeabilized in 0.25% Triton-X/PBS and then washed in PBS. Cells were treated with RNase A, RNase H, or no RNase control at 37 °C for 15 min. RNase inhibitors were added, and cells were cross-linked with formaldehyde. Cells were resuspended in lysis buffer (1% SDS, 50 mM Tris·HCl at pH 8, 10 mM EDTA, RNase inhibitors) and sonicated. Immunoprecipitation was performed using Dynabeads MyOne Streptavidin Beads (Thermo Fisher). Beads were incubated with the samples for 30 min and then washed five times in lysis buffer. Samples were eluted at 95 °C for 15 min in elution buffer (1% SDS, 10 mM EDTA, 0.1 mM $NAHCO_3$ at pH 8). Samples were then divided and treated with either RNase or DNase. Enrichment of DNA in RNase-treated samples was determined by qPCR, using the following primer sets: PTENP1_F/R, paRNA_F/R, PTENex1_F/R, PTENex9_F/R (Table S4). The RNA in DNase-treated samples was quantified by RT-qPCR, using the same primer sets as earlier. Scramble control was subtracted from the immunoprecipitate, and data were normalized to the sample input.

PTENpg1asRNA Exon1 and Truncated Variant Transcription. PTENpg1asRNA exon one and truncated variants (Table S2) were in vitro transcribed from linearized pcDNA 3.1 plasmids (containing the various inserts), using T7 RNA polymerase (19). RNA constructs were purified from transcription components by denaturing gel electrophoresis.

Electrophoretic Mobility Shift Assays. Protein lncRNA binding reactions were performed in a final volume of 35 μL and contained 637, 500, 700, or 600 ng PTENpg1asRNA exon1, F4R1, F5R2, or GFP RNA, respectively. All lncRNA variants were refolded by heating at 90 °C for 80 s before snap cooling on ice. Refolding was performed in a buffer containing 50 mM Hepes at pH 7.5, 150 mM KCl, 1.5 mH $MgCl_2$ 1 mM TCEP, and 20% glycerol for all lncRNAs. After refolding, protein was incubated with lncRNA for 30 min on ice. Dnmt3a/DnmtL-CD binding reactions contained between 1× and 10× molecular excess of the protein to RNA. DnmtL-CD binding reactions between 5× and 50× the molecular excess of the protein to RNA. Samples were loaded onto a 4.5% TBE acrylamide gel (containing 2.5 mM $MgCl_2$) and run at 250 V for 4 h at 4 °C. Gels were stained with SYBR gold (Invitrogen) and visualized using a Typhoon FLA 900 biomolecular imager.

ACKNOWLEDGMENTS. This project was supported by National Institute of Allergy and Infectious Diseases P01 AI099783-01, NIH DK104681-02, and ARC Future Fellow FT130100572 (to K.V.M.), and by the Swedish Research Council K2013-64X-20432-07-4, the Swedish Cancer Foundation 15 0768, Radiumhemmets Research Funds 144063, and the Swedish Childhood Cancer Foundation PR2015-0009 (to D.G.).

1. Steck PA, et al. (1997) Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat Genet* 15: 356–362.
2. Pei B, et al. (2012) The GENCODE pseudogene resource. *Genome Biol* 13:R51.
3. Puget N, et al. (2002) Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot. *Am J Hum Genet* 70:858–865.
4. Hawkins PG, Morris KV (2010) Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* 1:165–175.
5. Scarola M, et al. (2015) Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA. *Nat Commun* 6:7631.
6. Suo G, et al. (2005) Oct4 pseudogenes are transcribed in cancers. *Biochem Biophys Res Commun* 337:1047–1051.
7. Dahia PL, et al. (1998) A highly conserved processed PTEN pseudogene is located on chromosome band 9p21. *Oncogene* 16:2403–2406.
8. Teng DH, et al. (1997) MMAC1/PTEN mutations in primary tumor specimens and tumor cell lines. *Cancer Res* 57:5221–5225.
9. Poliseno L, et al. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465:1033–1038.
10. Johnsson P, et al. (2013) A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat Struct Mol Biol* 20:440–446.
11. Napoli S, Pastori C, Magistri M, Carbone GM, Catapano CV (2009) Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J* 28:1708–1719.
12. Han J, Kim D, Morris KV (2007) Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci USA* 104:12422–12427.
13. Roberts TC, Andaloussi SE, Morris KV, McClorey G, Wood MJ (2012) Small RNA-mediated epigenetic myostatin silencing. *Mol Ther Nucleic Acids* 1:e23.
14. Eckstein F (2000) Phosphorothioate oligodeoxynucleotides: What is their origin and what is unique about them? *Antisense Nucleic Acid Drug Dev* 10:117–121.
15. Di Ruscio A, et al. (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503:371–376.
16. Morris KV (2009) Long antisense non-coding RNAs function to direct epigenetic complexes that regulate transcription in human cells. *Epigenetics* 4:296–301.
17. Johnsson P, Lipovich L, Grander D, Morris KV (2014) Evolutionary conservation of long noncoding RNAs; sequence, structure, function. *Biochim Biophys Acta* 1840: 1063–1071.
18. Hawkins PG, Santoso S, Adams C, Anest V, Morris KV (2009) Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. *Nucleic Acids Res* 37:2984–2995.
19. Batey RT, Sagar MB, Doudna JA (2001) Structural and energetic analysis of RNA recognition by a universally conserved protein from the signal recognition particle. *J Mol Biol* 307:229–246.

Supporting Information

Lister et al. 10.1073/pnas.1621490114

SI Materials and Methods

ODN Transfections. Control (mirN367), ODN1, and ODN2 were transfected into cells to a final concentration of 100 nM using Lipofectamine 2000 (Life Technologies). ODNs were transfected into both 293HEK and the PTENpg1 asRNA exon1 overexpressing 293-HEK PG1 cells.

qRT-PCR. qRT-PCR analysis was performed using KAPA SYBR FAST qPCR Master Mix (KAPA Biosystems). Plate was placed in a ViiA 7 Real-Time PCR system (Life Technologies). Cycling conditions: 95 °C for 2 min and then 40 cycles of 95 °C for 3 s and 60 °C for 30 s.

Directional RT Analysis of Gene Expression. Directional reverse transcription (directional RT) was performed on HeLa total RNA, using the PTENproR1 primer (Table S4) to strand-specifically convert any RNA transcripts originating from the PTEN 5′ UTR into ssDNA. The resulting product was PCR amplified using primer set PTENproF2/R2 and run on a 2% agarose gel (Fig. 1*B*).

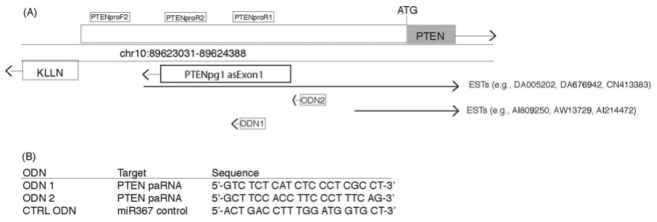
ChIP. ChIP analysis was carried out in 293HEK cells using anti-DNMT3a (Abcam, cat. no. ab2850). The ChIP was performed 48 h posttransfection with ODNs (Fig. 2 *D–I*) or various biotin-labeled truncations of the full-length PTENpg1 alpha exon 1 transcript (Fig. 3*A*) following previously described techniques (4–6). The relative enrichment of DNMT3a was determined at the PTEN promoter using primer sets PTENproF2/R2 (Fig. 1*F* and Table S4) or PTENproF1/R1 (Fig. 3*B* and Table S4). Any IgG or no antibody values are first subtracted from the resultant IP and input values, and then each sample is standardized relative to the sample input.

Dnmt3a/DnmtL-CD Purification. Dnmt3a (residues 284–910) and DnmtL-CD (residues 178–379) were cloned into a modified petDuet vector. Dnmt3a/DnmtL-CD proteins were coexpressed in *Escherichia coli* strain Rosetta 2 (DE3). The transformants were grown at 37 °C in LB medium and induced at an OD_{600} of 0.6 with IPTG and further incubated for 20–24 h at 18 °C. Dnmt3a/DnmtL-CD was purified from the supernatant of the cell lysate by three-step liquid chromatography. Nickel affinity, heparin affinity, and gel filtration chromatography were used and the purified protein complex stored in 50 mM Hepes at pH 7.5, 150 mM NaCl, 10% glycerol, and 1 mM TCEP. The purified proteins were estimated to be >90% pure by Coomassie blue-stained SDS/PAGE and were concentrated to 2–3 mg/mL for electrophoretic mobility shift assays. DnmtL-CD (residues 178–379) was cloned into pLIC-HK vector and overexpressed using the same methods as Dnmt3a/DnmtL-CD. DnmtL-CD was purified from cell lysate using Ni2+ affinity and gel filtration chromatography and stored in 50 mM Hepes at pH 7.5, 150 mM NaCl, 10% glycerol, and 1 mM TCEP. Purified DnmtL-CD was estimated to by >95% pure and was concentrated to 9–10 mg/mL.

T7-Transcribed Synthetic RNA Pulldown in Presence of ODNs. Synthetic biotinylated ncRNAs were generated by T7 transcription using the Ampliscribe T7-Flash Biotin –RNA Transcription Kit (Epicentre Biotechnologies) according to the manufacturer's instruction. Templates for T7 transcription were prepared by PCR of pcDNA3.1 plasmids expressing the relevant ncRNAs. Primers pcDNA3.1 F/R were used for PCR amplification (Table S4). PTENpg1 alpha exon 1 was in vitro transcribed using Durascribe, with biotin-linked dCTPs (described in detail in ref. 5). The resultant biotin PTENpg1 α exon 1 transcript was transfected into 293HEK cells at a concentration of 100 nM 18 h after transfection of either control ODN mirN367 or ODN2 transfection (*SI Materials and Methods*). Thirty hours after transfection of biotin conjugated transcripts, cells were cross-linked with formaldehyde at 1% final concentration for 10 min at room temperature followed by the addition of glycine to a final concentration of 0.125 M and a further incubation for 10 min at room temperature. Cells were then lysed with ChIP lysis buffer (5 mM Pipes, 85 mM KCl, and 0.5% Nonidet P-40) supplemented with PMSF on ice for 20 min. Chromatin was sheared by sonication. Cell lysates containing sheared chromatin were incubated with Dynabeads M280 Streptavidin (Life Technologies) prepared according to the manufacturer's instructions for 1 h on a rotating platform. Beads were pulled down with a magnet for 3 min and washed with low salt immune complex wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris·HCl at pH 8.1, 150 mM NaCl), high salt immune complex wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris·HCl at pH 8.1, 500 mM NaCl), LiCl Immune complex wash buffer (0.25 M LiCl, 1% Nonidet P-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris·HCl at pH 8.1), and TE buffer (10 mM Tris·HCl, 1 mM EDTA at pH 8.0). Each wash step was carried out for 3 min on a rotating platform. Streptavidin bead-biotinylated RNA–DNA complexes were resuspended in Elution buffer (1% SDS, 0.1 M $NaHCO_3$) and heated at 95 °C for 5 min to dissociate biotin RNA from beads. Eluted biotinylated RNA complexes were removed from beads via magnet pull-down, and supernatants were analyzed by qPCR for enrichment at the PTEN promoter. Samples were then DNase, RNase A, and proteinase K treated so as to remove everything from elutes except dsRNAs. These were reverse transcribed and analyzed by qPCR for PTEN promoter enrichment. Primer sets used for analyzing the PTEN promoter were PTENproF1/R1, PTENproF2/R2, and PTENproF2/R1 (Fig. 2*C* and Table S4).

Truncated PTEN Alpha Exon 1 and F4R1 Deletions ChIP. Various truncated versions of PTEN alpha exon 1 (Table S2) and mutants of F4R1 (Table S3) were generated to be expressed from the CMV promoter in the context of pcDNA3.1 (Genewiz). The constructs (Tables S2 and S3) were in vitro transcribed using Durascribe with biotin-linked dCTPs (described in detail in ref. 5). The resultant biotin-labeled transcripts and a GFP-biotin control were transfected into 293HEK cells (50 nM) and assessed by RIP 48 h later using PTENproF2/R2 (Table S4) primers for detection at the PTEN promoter.

PTEN Promoter CpG Methylation Post-ODN2 Treatment. Genomic DNA was isolated from Hek293 cells after 48 h treatment with ODN scrambled control and ODN2 (Table S4). Briefly, 200 ng DNA was digested with the methyl-cytosine-specific restriction enzyme McrBc (New England Biolabs) overnight at 37 °C. Twenty-four hours later, the enzyme was heat inactivated at 65 °C for 1 h and qRT-PCR performed (Kapa Biosystems), using Killin F/R primers (Table S4), and standardized to uncut input. The delta CT values were converted to fold-change values, and the ratio between ODN2/ODN scrambled control treated cells was calculated.



(B)

ODN	Target	Sequence
ODN 1	PTEN paRNA	5'-GTC TCT CAT CTC CCT CGC CT-3'
ODN 2	PTEN paRNA	5'-GCT TCC ACC TTC CCT TTC AG-3'
CTRL ODN	miR367 control	5'-ACT GAC CTT TGG ATG GTG CT-3'

Fig. S1. Location of directional RT-PCR primers and antisense ODNs relative to the PTEN promoter. (*A*) A schematic of the PTEN promoter is shown depicting the PTENpg1 asRNA exon1 binding site and location of antisense ODNs, as well as RT-PCR primers used in directional RT analysis of the PTEN-associated paRNA. (*B*) The sequence of the ODNs and control ODN (miRN367, ref. 1) used in the presented study.

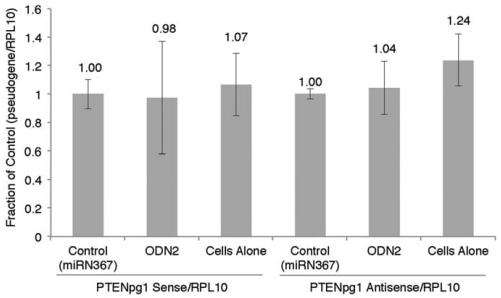


Fig. S2. PTENpg1 and PTENpg1 asRNA expression after ODN2 treatment. The 293HEK cells were transfected in triplicate with ODN2, control ODN miRN37 target site (1, 2). The averages of triplicate-treated cultures are shown with the SDs.

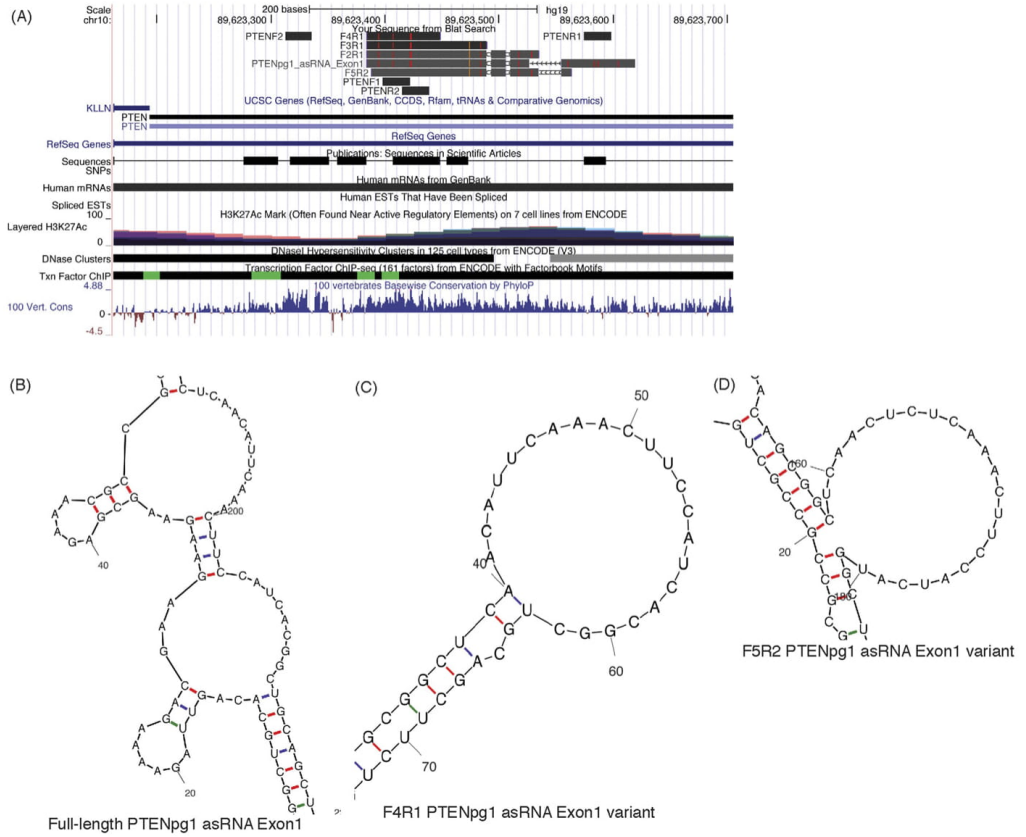


Fig. S3. Analysis of the PTENpg1 asRNA alpha exon 1 and various truncations of PTENpg1 asRNA exon1 target sites in the PTEN promoter. (*A*) A screen shot from UCSC genome browser is shown depicting the PTEN promoter and homologous target sites for PTENpg1 asRNA exon1 and the various truncations of PTENpg1 asRNAs assessed. (*B–D*) The predicted DNMT3a binding bulge in both (*B*) full-length and (*C*) F4R1 variant of PTENpg1 asRNA exon 1, and lost in (*D*) F5R2 variant, as determined from M-Fold. The conserved sequence that appears required for DNMT3a interactions with PTENpg1 asRNA exon 1 at the PTEN promoter (ACAUUCAAAC UUCCAUCACGGC) is shown in PTENpg1 asRNA exon 1 and F4R1, but lost in F5R2.

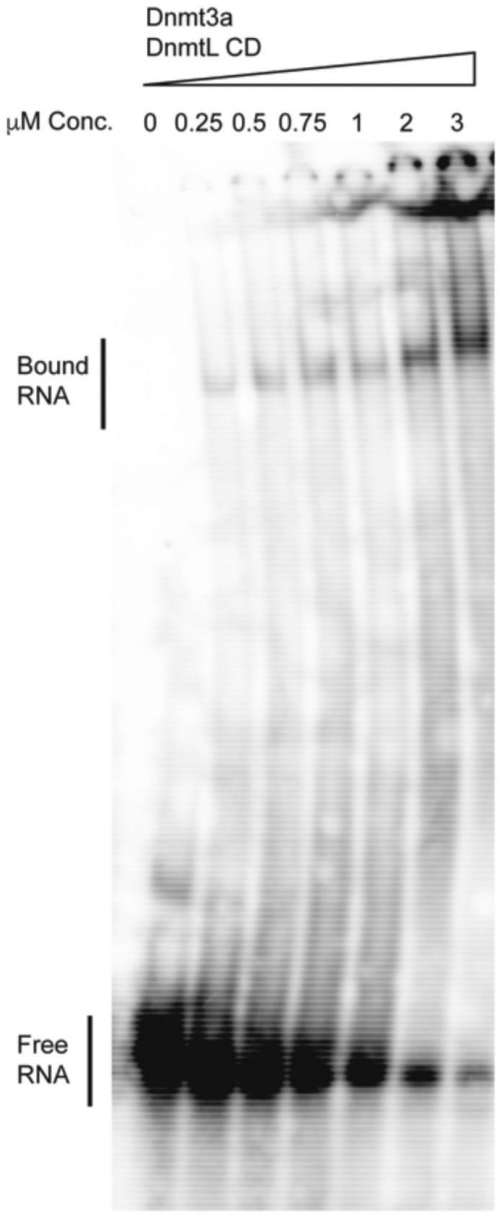


Fig. S4. EMSA analysis of PTENpg1 asRNA truncated F4R1 binding to DNMT3a in the presence of DNMTL-CD. Protein concentration is indicated above, and the free RNA and bound RNA-DNMT3a are indicated.

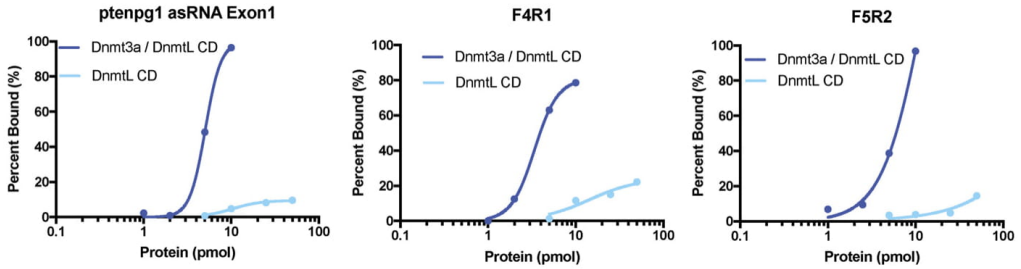


Fig. S5. Analysis of the bound fraction of EMSA for PTENpg1 asRNA exon 1, F4R1 and F5R2 binding to Dnmt3a in the presence of DnmtL-CD. Bands were quantified using ImageJ and plotted in a log scale with a curve fit using Prism 7.

Table S1. Reported EST BG772190 found on chr10:89623172–89623881

Construct name	Long noncoding RNA genetic sequence
EST BG772190	5′TACCGCCCCCTGCCCTGCCCTGCCCTCCCCTCGCCCGGCGCGGTCCCGTCCGCCTCTCGCTCGCCTCCCGCCTCCCCTCGGTCTTCCGAGG-CGCCCGGGCTCCCGGCGCGGCGGCGGAGGGGGCGGGCAGGCCGGCGGGCGGTGATGTGGCGGGACTCTTTATGCGCTGCGGCAGGATACGC-GCTCGGCGCTGGGACGCGACTGCGCTCAGTTCTCTCCTCTCGGAAGCTGCAGCCATGATGGAAGTTTGAGAGTTGAGCCGCTGTGAGGCGA-GGCCGGGCTCAGGCGAGGGAGATGAGAGACGGCGGCGGCCCGCGGCCCGGAGCCCCTCTCAGCGCCTGTGAGCAGCCGCGGGGGCAGCGCCC-TCGGGGAGCCGGCCGGCCTGCGGCGGCGGCAGCGGCGGCGTTTCTCGCCTCCTCTTCGTCTTTTCTAACCGTGCAGCCTCTTCCTCGGCTT-CTCCTGAAAGGGAAGGTGGAAGCCGTGGGCTCGGGCGGGAGCCGGCTGAGGCGCGGCGGCGGCGGCGGCACCTCCCGCTCCTGGAGCGGaG-GGGAGAAGCGGCGGCGGCGGCCGCGGCtGGCTGCAGCTCCAGGGAGGGGGTCTGAGTCGCCTGTCACCATTTCCAGGGCTGGGAACGC-CGGAGAGTcGGTCTCTCCCCTTCTACTGnCTCCAACACGGCCGGCGGCTGGCggcggcaACATCCAGGGACCCGGG3′

Table S2. Truncated PTEN alpha exon 1 variants

Construct name	Long noncoding RNA genetic sequence
PTENpg1 antisense alpha exon 1 (described in ref. 3)	5′GGAAGAGGCTGCACAGTTAGAAAAGACGAAGAAGAAGCGAGAAACGCCGCCGCTGCCGGCGCCGCCCCCGCGGATGCT-CACGGGTTGCTGAGAGGGGCTTCAGGCCGGGCCGGGCCGGGCCGCCGCCGCCGCCGCTCTCTGTCTCTCATCTCC-CTCGCCTGAGCCCGGCCTCGCCTCACAGCGGCTCAACATTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGGAGAG-CATCTTCTAG3′
F2R1 PCR set 2 (variant 1 PTENpg1asRNA alpha, 220 bp fragment)	5′CCGGCGCCGCCCCCGCGGATGCTCACGGGTTGCTGAGAGGGGCTTCAGGCCGGGCCGGGCCGGGCCGCCGCCGCC-GCCGCCGTCTCTGTCTCTCATCTCCCTCGCCTGAGCCCGGCCTCGCCTCACAGCGGCTCAACATTCAAACTTC-CATCACGGCTGCAGCTTCTGAGAGGAGAGCATCTTCTAGTTTTT3′
F3R1 PCR set 3 (variant 2 PTENpg1asRNA alpha, 164 bp fragment)	5′GGGCCGGGCCGCCGCCGCCGCCGCCGTCTCTGTCTCTCATCTCCCTCGCCTGAGCCCGGCCTCGCCTCACAGCGGC-TCAACATTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGGAGAGCATCTTCTAGTTTTT3′
F4R1 PCR set 4 (variant 2 PTENpg1asRNA alpha, 107 bp fragment)	5′GGGTAGTGAAGGCTAGCGGGCCTCGCCTCACAGCGGCTCAACATTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGG-AGAGCATCTTCTAGTTTTTGGTACCTATCTGA3′
F5R2 PCR set 4 (variant 2 PTENpg1asRNA alpha, 217 bp fragment)	5′TAGTGAAGGCTAGCGCGCCGCCGCTGCCGGCGCCGCCCCCGCGGATGCTCACGGGTTGCTGAGAGGGGCTTCAGG-CCGGGCCGGGCCGGGCCGCCGCCGCCGCCGTCTCTGTCTCTCATCTCCCTCGCCTGAGCCCGGCCTCGCCTC-ACAGCGGCTCAACTCTCAAACTTCCATC**AT**GGCTGCAGCTTCCGAGAGGTTTTTGGTACCTATCTGA3′

The experimentally determined 3a binding bulge is underlined in the F4R1 and full-length PTENpg1 alpha exon 1 and the AT insertion in bold that disrupts the 3a binding bulge for the F5R2 variant. Underlined segments represent predicted loop promoter-associated RNA-binding loci.

Table S3. The F4R1 functional domains

F4R1 variants	Long noncoding RNA genetic sequence
F4R1	5′GGGTAGTGAAGGCTAGCGGGCCTCGCCTCACAGCGGCTCAACATTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGGAGAGCATCTTCTAGTTTTTG-GTACCTATCTGA3′
C1 Construct	5′TAGTGAAGGCTAGCGGGCCTCGCCTCACAGCGGCTCACGAATGCAGCTTCTGAGAGGAGAGCATCTTCTAGTTTTTGGTACCTATCTGA3′
C2 Construct	5′CTCGCCTCACAGCGGCTCAACATTCAAACTTCCATCACGGCTGCAGCTTCTGAGAGGAGAGC3′
C3 Construct	5′GGGTAGTGAAGGCTAGCGGGCGAATCTTCTAGTTTTTGGTACCTATCTGA3′

Several variations of the F4R1 transcript were derived. Deletion construct C1 contains a change to the major loop of F4R1, resulting in the replacement of the major loop with another loop sequence. Construct C2 contains a deletion of "lower" stem and construct C3 contains a deletion of the "upper" stem. Underlined segments represent predicted loop promoter-associated RNA-binding loci.

Table S4. Primers and ODN sequences used in analysis (5′-3′ shown)

Name	Long noncoding RNA genetic sequence
PTENproF1	GCTGCAGCCATGATGGAAGTTTGA
PTENproR1	AAAGACGAAGAGGAGGCGAGAAAC
PTENproF2	TGATGTGGCGGGACTCTTTATGC
PTENproR2	TCACAGCGGCTCAACTCTCAAACT
PTENF3	AGAAAGCTTACAGTTGGGCCCTGT
PTENR3	GCCACAGCAAAGAATGGTGATGCT
PTENP1ex1_F	GGAAGAGGCTGCACAGTTA
PTENP1ex1_R	CTAGAAGATGCTCTCCTCTCA
paRNA_F	ATGTGGCGGGACTCTTTATG
paRNA_R	GCGGCTCAACTCTCAAACT
PTENex1_F	TGCCATCTCTCTCCTCCTT
PTENex1_R	CGAATCCATCCTCTTGATATCTCC
PTENex9_F	TGTAATCAAGGCCAGTGCTAAA
PTENex9_R	AGCATCCACAGCAGGTATTATG
PTEN unspliced F	AAAGCTGGAAAGGGACGAACTGGT
PTEN unspliced R	TCTCAGATCCAGGAAGAGGAAAGG
mirN367 (control) ODN	GTGTGGGGTTTTAGCTTCGTGAA
ODN1	GTCTCTCATCTCCCTCGCCT
ODN2	GCTTCCACCTTCCCTTTCAG
Scrambled ODN2	CTAACTCTCCGTGTCCTCCT
pcDNA3.1 F	CCCACTGCTTACTGGCTTATC
pcDNA3.1 R	CAGATGGCTGGCAACTAGAA
Killin F	ACACAAGCACCCACATCCAAA
Killin R	AGTCCTTTGGCTTGCTCTTAG

7.3 Journal Copyright Information

Below are copyright excerpts taken directly from the websites of the journals in which chapters of this thesis were published.

For paper 1 in Chapter 2. Taken directly from the RSOS journal FAQs page:

http://rsos.royalsocietypublishing.org/faq#question15

"How do I seek permission to reproduce material from Royal Society Open Science?

Papers in *Royal Society Open Science* are published under a Creative Commons Attribution License which allows reuse of material with attribution, therefore no permission is required from the authors/publisher so long as the original authors and source are cited."

For paper 2 in Chapter 3. Taken directly from the PNAS journal licenses page:

http://www.pnas.org/page/authors/licenses

"Except, as noted above, for open access articles submitted beginning September 2017, authors retain copyright, but grant to PNAS an exclusive License to Publish. Authors and their employing institution or company retain extensive rights for use of their materials after publication in PNAS and for intellectual property. These rights are retained without requiring explicit permission from PNAS, provided the full journal reference is cited."

"Authors retain the following rights:

…The right to include the article in the author's thesis or dissertation."