TRANSCRIPTIONAL GENE SILENCING IN MAMMALIAN

CELLS BY MICRO-RNAS THAT TARGET GENE PROMOTERS

APPROVED BY SUPERVISORY COMMITTEE

_____

David R. Corey, Ph.D., Advisor

_____

Michael A. White, Ph.D., Chairman

_____

Eric N. Olson, Ph.D.

_____

Alexander Pertsemlidis, Ph.D.

ACKNOWLEDGEMENTS

I would like to thank my family, especially my mother and father, for their support throughout my academic pursuits as well as throughout my life. I would also like to thank my best friends and colleagues, Drs. Brian and Alison Erickson, for being such a major part of both my academic development and my life experience.

I would like to thank all the current and previous members of the Corey laboratory. Specifically, I would like to thank Dr. Randall Beane and Dr. Jacob Schwartz for their camaraderie during our graduate careers together.

I would like to thank all the members of my dissertation committee: Dr. Michael White, Dr. Eric Olson, and Dr. Alexander Pertsemlidis. In addition, I would like to thank the chairman of the Cell Regulation graduate program, Dr. Paul Sternweis. Each of them played an integral role in my development as a scientist.

Finally, I would like to thank my graduate advisor, Dr. David Corey, for all of the opportunities given to me during my tenure as a student in his laboratory. The experience and training I received while in his laboratory have been an invaluable part of my preparation for a career in science.

TRANSCRIPTIONAL GENE SILENCING IN MAMMALIAN CELLS BY MICRO-RNAS THAT TARGET GENE PROMOTERS

by

Scott T. Younger

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

April 2011

Copyright

by

Scott T. Younger, 2011

All Rights Reserved

TRANSCRIPTIONAL GENE SILENCING IN MAMMALIAN

CELLS BY MICRO-RNAS THAT TARGET GENE PROMOTERS

Scott T. Younger, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2011

David R. Corey, Ph.D.

A rich history exists for RNA-based regulation of gene transcription. It was reported more than a decade ago that RNA is capable of inducing DNA methylation and transcriptional gene silencing in plants. It was subsequently shown that small RNAs are involved in the establishment of heterochromatic regions of the yeast genome. More recently it has been demonstrated that small duplex RNAs designed to be complementary to gene promoters are potent regulators of gene transcription in mammalian cells.

Potent and robust transcriptional regulation by designed small RNAs suggests the existence of endogenous mechanisms that facilitate recognition of gene promoters by small RNAs in mammalian cells. microRNAs (miRNAs) are endogenous small RNAs that regulate gene expression post-transcriptionally through base complementarity to

target sequences within 3'-UTRs of mRNA transcripts. In this body of work I test the hypothesis that miRNAs can also recognize sequences within gene promoters using two alternative approaches.

In the first approach I computationally evaluate the potential for miRNAs to recognize gene promoters by performing a genome-wide survey of putative miRNA target sites within promoter sequences. In the second approach I use the well-characterized human progesterone receptor (PR) gene as a model to experimentally validate that miRNAs possess the ability to regulate transcription in a cell culture system.

I have found that gene promoters are significantly enriched for miRNA target sites. Furthermore, the frequency of miRNA target sites within promoter sequences is comparable to their frequency within 3'-UTRs. I experimentally screened multiple miRNAs predicted to target the PR gene promoter, identified several that were capable of inhibiting transcription of the PR gene, and characterized the mechanism of transcriptional silencing.

miRNAs have been understood to regulate gene expression at the post-transcriptional level through recognition of 3'-UTRs within mRNA transcripts. My study extends miRNA function to recognition of sequences within gene promoters. Sequence specific recognition of gene promoters by miRNAs may complement protein transcription factors. In addition, the ability of small RNAs to rapidly evolve specificity for new sequences would have evolutionary advantages.

TABLE OF CONTENTS

PUBLICATIONS

1. Janowski BA, **Younger ST**, Hardy DB, Ram R, Huffman KE, Corey DR. "Activating gene expression in mammalian cells with promoter-targeted duplex RNAs". Nat. Chem. Biol. 2007 Mar;3(3):166-73.

2. Schwartz JS, **Younger ST**, Nguyen N, Hardy DB, Monia BP, Corey DR, Janowski BA. "Antisense Transcripts are Targets for Activating Small RNAs". Nat. Struct. Mol. Biol. 2008 Aug;15(8):842-8.

3. **Younger ST**, Corey DR. "The Puzzle of RNAs that Target Gene Promoters". Chembiochem. 2009 May 4;10(7):1135-9.

4. **Younger ST**, Pertsemlidis A, Corey DR. "Predicting potential miRNA target sites within gene promoters". Bioorg. Med. Chem. Lett. 2009 Jul 15;19(14):3791-4.

5. Yue X, Schwartz JC, Chu Y, **Younger ST**, Gagnon K, Elbashir S, Janowski BA, Corey, DR. "Regulation of Transcription by Small RNAs Complementary to Sequences Downstream from the 3' Termini of Genes". Nat Chem Biol. 2010 Aug;6(8):621-9.

6. Chu Y, Yue X, **Younger ST**, Janowski BA, Corey DR. "Involvement of Argonaute Proteins in GeneSilencing and Activation by RNAs that target a Noncoding Transcript at the Progesterone Receptor Promoter". Nucleic Acids Res. 2010 Nov 1;38(21):7736-48.

7. **Younger ST**, Corey DR. "Transcriptional Gene Silencing in Mammalian Cells by miRNA Mimics That Target Gene Promoters". Nucleic Acids Res. 2011 In Press.

8. **Younger ST**, Corey DR. "Transcriptional Regulation by miRNA Mimics That Target Sequences Downstream of Gene Termini". In Review.

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDS	coding sequence
ChIP	chromatin immunoprecipitation
DNA	deoxyribonucleic acid
hnRNA	heteronuclear RNA
MFE	minimum free energy
miRNA	microRNA
mRNA	messenger RNA
MSP	methylation specific PCR
ncRNA	noncoding RNA
PCR	polymerase chain reaction
PR	progesterone receptor
qPCR	quantitative PCR
RIP	RNA immunoprecipitation
RNA	ribonucleic acid
siRNA	small interfering RNA
TSS	transcription start site
UTR	untranslated region

Chapter 1: Introduction

1.1 RNA interference

While characterizing the ability of antisense RNA transcripts to interfere with gene activity, Fire *et al.* made a striking observation. In the study, 742 base RNA fragments that were complementary to the *unc-22* gene, which encodes a myofilament protein required for sustaining muscle contraction, were directly injected into *C. elegans*[1-4]. While neither sense nor antisense oriented RNA fragments alone were effective inhibitors of gene activity, introduction of both sense and antisense RNA fragments at the same time resulted in a strong twitching phenotype that was propagated from the injected animal to its progeny[4]. The authors concluded that the interfering activity was specific to double-stranded RNA and that the process likely involved some type of catalytic component. The phenomenon of gene inhibition by double-stranded RNA has been termed RNA-interference (RNAi).

1.2 Mechanism of RNAi

Since the discovery of RNAi in 1998 a significant effort has been focused on characterizing the mechanism of action by which double-stranded RNA interferes with gene expression. These efforts have been aided by data that had been previously acquired from many years of experiments in plants. Reports as early as 1990 found that transgene expression in plants could result in coordinate suppression (co-suppression) of both the introduced transgene and the homologous endogenous gene[5-7]. One initial explanation for this result was that full length antisense transcripts might be produced

from the transgene and that these antisense transcripts were guiding degradation of the sense-oriented mRNA. Northern blot experiments revealed, however, that co-suppression was accompanied not by full length antisense RNA transcripts, but by short 25 nucleotide RNA fragments in both the sense and antisense orientations[8].

At the same time a genetic screen in *C. elegans* identified *rde-1*, a member of the Argonaute (AGO) gene family which is conserved from plants to humans, as being required for RNAi[9]. Shortly afterwards, independent experiments using biochemical purification approaches in *Drosophila* cells linked these findings by demonstrating that the enzymatic production of small RNAs associated with gene silencing co-fractionates with an RNA-directed nuclease identified to be an AGO family member[10-11]. Subsequent *in vitro* experiments revealed that the sequence-specific mediators of RNAi are 21-22 nucleotide RNA fragments termed short interfering RNAs (siRNAs) that guide AGO proteins to complementary RNA sequences and result in the cleavage of the targeted RNA[12].

When introduced into worms or insects, long double-stranded RNAs are processed into siRNAs that are potent silencers of gene expression. In mammalian cells, however, double-stranded RNAs greater than 30 bp in length invoke a double-stranded RNA-induced interferon response[13-15]. The characterization of chemically synthesized 21 bp siRNAs that silenced gene expression in *Drosophila* cells suggested that a similar approach might be successful in mammalian model systems[12]. Indeed, it was soon after shown that chemically synthesized 21 bp siRNAs are capable of targeting complementary mRNA sequences and inhibiting gene expression in cultured mammalian cells without inducing an interferon response[16]. The ease of synthesis of short RNA sequences coupled

with advances in small RNA delivery into cultured cells have made siRNAs ubiquitous tools for studying gene function. Furthermore, their ability to potently inhibit gene expression has made siRNAs promising candidates for the development of therapeutic agents.

1.3 RNA-induced transcriptional gene silencing

The characterization of post-transcriptional gene silencing by double-stranded RNA helped explain previous observations of DNA methylation that had been made in plants. Transgene-induced DNA methylation was detected as early as 1989 in tobacco plants during experiments that involved two sequential transformations of DNA plasmids. In these studies, introduction of a second transgene often resulted in suppression of both transgenic constructs and inhibition was accompanied by DNA methylation at sequences that were present in both vectors[17-18]. A similar result was described in 1994 using viroid cDNA units and implicated a role for RNA in directing DNA methylation of homologous sequences[19]. In 2000, northern blot experiments in *Arabidopsis* revealed that DNA methylation by RNA is triggered by double-stranded RNAs that ultimately get processed into ~23 nucleotide small RNAs[20].

A direct role for the involvement of the RNAi machinery in regulating gene expression at the level of DNA, specifically heterochromatin formation, was observed in the fission yeast *S. pombe* in 2002. Deletion of the RNAi machinery resulted in de-repression of heterochromatic regions of the yeast genome[21-22]. Subsequent biochemical purification of heterochromatin-associated protein Chp1 revealed a physical association between the single *S. pombe* AGO homolog and other protein factors which

collectively became termed the RNA-induced initiation of transcriptional silencing (RITS) complex[23]. Around the same time, a requirement for the RNAi machinery in heterochromatin formation was also reported in *Drosophila* [24-25]. It was becoming clear that small RNAs had a remarkable potential for regulating gene expression at both the transcriptional and the post-transcriptional levels.

1.4 Promoter-targeting small duplex RNAs

Transcriptional regulation by small RNAs in plants, yeast, and flies prompted the question of whether small RNAs were also capable of regulating transcription in mammalian cells. In 2004 a report was published claiming that synthetic small duplex RNAs designed to be fully complementary to the promoter region of a transduced reporter construct were capable of silencing expression of the transgene in human cells[26]. Gene silencing was observed at the level of transcription and was associated with DNA methylation of the transgene promoter. Shortly afterwards it was shown that designed small duplex RNAs could also silence transcription of endogenous genes in human cells, but in the absence of DNA methylation[27-28].

Mechanistic studies in the following years revealed that small RNA-induced transcriptional gene silencing in humans, like that in other organisms, is mediated by AGO proteins[29-30]. The human genome encodes four AGO family members (AGO1-4) and convincing evidence points to AGO2 as being a key mediator of transcriptional silencing[29,31]. However, some reports have suggested a requirement for AGO1[30,32]. Most of the studies on small RNA-induced transcriptional gene silencing in humans have used different model genes and it's possible that the genomic context of the target promoter

dictates which AGO isoform is used. In either case, promoter-targeting small RNAs silence transcription of the downstream gene and are associated with repressive histone modifications at the targeted promoter[30,32-33].

While promoter-targeting small RNAs affect gene expression at the level of the chromosome, the molecular target of these RNAs was initially unclear. AGO proteins are well-characterized for their role in mediating small RNA recognition of RNA transcripts, but no evidence exists for AGO-mediated recognition of DNA in human cells. Later reports revealed that promoter-targeting small RNAs actually recognizing RNA transcripts that are produced from within the targeted promoter[32,34]. In one case this RNA was a rare variant of the target gene mRNA that was initiating from an upstream promoter[32]. In another, the RNA was an antisense transcript that initiated from within the gene and was transcribed through the promoter region[34]. In both cases the production of the RNA species that overlapped the targeted promoter was required for transcriptional silencing. The observation of RNA being produced from gene promoters highlights the previously unappreciated complexity of transcription occurring within the human genome.

1.5 The transcriptome

After the initial sequencing of the human genome, the next step was to study its expression and regulation[35-36]. In the early 2000's the technical advances being made with hybridization arrays made this a feasible task. Several groups began using tiling arrays to study the transcriptional activity of individual chromosomes[37-39]. These studies all found that transcription was not restricted to protein coding regions, but was also

occurring within introns, promoters, and additional intergenic regions. As technology evolved and throughput increased, it became evident that while less than 2% of the human genome encodes protein, greater than 90% of the genome is transcribed into RNA[40-42]. The collection of all RNA transcripts produced within a cell has been termed the transcriptome.

Transcriptome surveys have revolutionized our understanding of the human genome's regulation and they remain a rich source for discovery[43-45]. They have revealed complex networks of noncoding RNA (ncRNA) transcripts that are produced from many regions of the genome. Strong evidence suggests that many ncRNAs regulate gene expression through the recruitment of chromatin-modifying complexes to specific genomic loci[46-48]. For some ncRNAs, the physical act of transcription modifies the local chromatin in a manner that allows for robust gene activation in response to stimuli[49]. Interestingly, it appears that a significant number of promoters for protein-coding genes in the human genome are overlapped by ncRNA transcripts[50]. This suggests that many genes have the potential to be transcriptionally regulated by small RNAs that are targeted to their promoter regions.

1.6 Small duplex RNAs that target gene termini

Transcriptome analyses have found that, in addition to promoter regions, many protein-coding genes are overlapped at their 3'-terminus by ncRNA transcripts[50-51]. The function of these RNAs remains unclear, but their properties are reminiscent of promoter-overlapping RNAs. In 2010 it was found that small duplex RNAs designed to be fully complementary to ncRNA transcripts that overlap gene termini were capable of

silencing gene expression in human cells[33]. The observed gene silencing was associated with the recruitment of AGO2 to the targeted ncRNA transcript and reduction of RNA Pol II occupancy on the promoter of the upstream gene.

In many cases the 5' and 3' ends of genes are separated by large genomic distances, making it difficult to imagine how small RNAs targeted to the terminus of a gene could affect activity at its promoter. Chromatin conformation capture experiments revealed, however, that the 5' and 3' ends of the genes targeted in the previously described study were in close physical proximity[33,52]. The physical proximity of gene promoters and termini appears to be a general phenomenon and has been termed gene looping[52-56].

Several reports have attributed important biological roles to gene looping[54,57-62]. Looping is commonly observed at actively transcribed genes and may play a role in transcriptional memory[52-53,62]. Transcriptional memory allows actively transcribed genes to be rapidly reactivated with kinetics much faster than genes undergoing initial activation. Furthermore, enhancer elements have been identified at the 3' end of the genes in which they regulate[54]. These studies highlight important functional roles for the crosstalk between gene promoters and termini, providing a potential model for how small RNA interactions at the 3' end of a gene could influence transcription.

1.7 microRNAs

In the initial report of RNAi it was suggested that the mechanism likely had some role in physiological gene silencing[4]. Five years prior to the discovery of RNAi, two reports characterized a small RNA gene (*lin-4*) that temporally inhibits expression of an mRNA

transcript (*lin-14*) through complementarity to its 3'-UTR in *C. elegans*[63-64]. At the time of their publication it was assumed that the small RNA gene was acting through simple antisense base pairing and that this small RNA was an anomaly of the *C.elegans* genome. The biological role of small RNAs in gene silencing was largely overlooked until 2000 when a second small RNA gene in *C. elegans* was identified[65]. This second small RNA (*let-7*) was of particular interest because its sequence and expression were highly conserved from nematodes to humans[66]. In the following year several reports characterized the existence of large numbers of small RNA genes in *C. elegans*, *D. melanogaster*, and humans[67-69]. The members of this new class of small RNA genes were termed microRNAs (miRNAs).

Upon their discovery, it was presumed that many miRNAs would play important physiological roles as post-transcriptional regulators of gene expression. This was based on the fact that both of the initially identified miRNAs, *lin-4* and *let-7*, were discovered through their involvement in *C. elegans* larval development[63-65]. Indeed, several reports in the following years described roles for miRNAs in a variety of processes from cell proliferation in *Drosophila* to hematopoietic lineage differentiation in mammals[70-71]. Mouse models have further demonstrated an integral role for miRNAs in development and disease[72-74].

miRNAs are produced from RNA transcripts that range in length from 10's to 100's of nucleotides long termed pri-miRNAs that contain one or more stem loop structures. The RNase III endonuclease Drosha excises individual ~60-70 base stem loop fragments called pre-miRNAs[75-77]. Pre-miRNAs are further processed by the RNase III enzyme Dicer into siRNA-like duplexes[78-80]. One strand of the duplex, termed the mature

miRNA, is ultimately loaded into an AGO protein which is the core component of the RNA-induced silencing complex (RISC)[10,12,16,81-83]. RISC is recruited to mRNA transcripts that are complementary to the miRNA and recognition results in the inhibition of gene expression.

A majority of miRNAs are expressed from genomic loci that are distant from protein-coding genes, suggesting that they are independently regulated[67-69]. However, examples do exist of miRNAs that are encoded within introns of protein-coding genes[84-86]. In these cases, expression of the miRNA appears to be coordinated with expression of the surrounding gene. Most miRNAs have highly regulated expression patterns. For example, *lin-4* and *let-7* have stage-specific expression patterns during larval development in *C. elegans*[66,68,85]. In humans, expression of many miRNAs is restricted to specific tissue types[69,71,87].

Following the discovery of miRNAs, several additional classes of endogenous small RNAs have been characterized. However, the biogenesis of these RNAs differs from that of miRNAs. For example, transcription through inverted repeats can result in the generation of siRNAs[88-90]. Overlapping transcripts that are produced from both strands of a DNA sequence can hybridize and subsequently generate siRNAs[91-93]. Piwi-interacting RNAs (piRNAs) are endogenous siRNAs that are derived from repetitive regions of the genome and also from transposons, however these siRNAs are specifically expressed in the germline of insects and mammals[94-100]. Most of these additional classes of small RNAs seem to be involved in the silencing of retrotransposons to maintain genome integrity as opposed to regulating the expression of specific genes like miRNAs[101-103].

1.8 miRNA targets

Determining the targets of miRNAs is a difficult task because most identified miRNAs do not have extensive complementarity to known mRNA transcripts. Despite widespread computational and experimental efforts to identify the specific determinants of miRNA target recognition, to date there are few steadfast rules[104-111]. Collectively, the previously mentioned studies have shed light on some recurring themes in miRNA target recognition. Perfect complementarity between bases 2-8 of the miRNA (termed the seed sequence) and its target mRNA are required for recognition[104,106]. In addition, the seed sequence target site of functional miRNAs tends to be conserved in orthologous mRNA transcripts in other species[107]. These studies have also revealed that single miRNAs have the potential to target multiple genes (multiplicity) and that multiple miRNAs can target a common mRNA (cooperativity)[112-113].

1.9 Mechanism of gene inhibition by miRNAs

The mechanism by which miRNAs inhibit gene expression remains a topic of debate. The initial reports on *lin-4* suggested that the miRNA was repressing translation of the target *lin-14* mRNA[64,114]. This observation resulted in the theory that small RNAs that are fully complementary to their mRNA targets (siRNAs) act by cleaving the mRNA whereas small RNAs with partial complementary to their mRNA targets (miRNAs) act by repressing translation of the mRNA. As it turns out, this theory is only partially correct.

With the emergence of more sensitive techniques for RNA detection it has been shown that *lin-4* does affect expression of its target *lin-14* mRNA[115]. Using microarray

technology it was subsequently shown that miRNAs decrease mRNA expression levels for many of their predicted target genes[113,116-118]. In rare cases, when complementarity between a miRNA and its target is near perfect, this decrease can be a result of cleavage of the mRNA target[119-120]. In most cases, however, it is likely that recognition of the target mRNA results in deadenylation, decapping, followed by degradation of the mRNA through canonical mRNA-turnover pathways[117,121-123]. This hypothesis has been strengthened by high-throughput analyses evaluating the effect of individual miRNAs on overall mRNA and protein expression[124-126].

1.10 Nuclear functions of miRNAs

Most studies on miRNAs have focused on their ability to inhibit gene expression through recognition of mRNA transcripts within the cytoplasm. However, multiple reports have described the presence of mature miRNA species within the nucleus[127-130]. Some miRNAs are exclusively localized to the nuclear compartment suggesting that they have functional roles that are distinct from the canonical miRNA pathway[127]. In addition to miRNAs, the AGO proteins that mediate miRNA function are also present in the nucleus[131-135]. AGO proteins are actively transported into the nucleus via specific importins, further supporting the idea that their nuclear presence is biologically relevant[135]. The function of nuclear miRNAs is not clear, but it is possible that they could direct transcriptional gene silencing in a manner similar to designed small duplex RNAs.

1.11 miRNA-induced transcriptional gene silencing

There have been isolated reports suggesting a potential for miRNA-induced transcriptional modulation in human cells[136-137]. One study found that introduction of an exogenous miRNA mimic resulted in the activation of genes that contain promoter regions with complementarity to the miRNA[136]. The specific miRNA was chosen based on its complementarity to the E-cadherin gene promoter, which had previously been targeted with designed small duplex RNAs[27,138]. Another report claimed that a miRNA was capable of targeting its own genomic location *in cis* and silencing transcription of the immediately adjacent gene[137]. The conclusions derived from that study are restricted to miRNAs that are encoded within promoter regions, of which 10 have been identified in the human genome (although this number is subject to change with the growing number of characterized miRNAs)[137].

1.12 Hypothesis

The observation that synthetic small duplex RNAs are capable of regulating gene transcription coupled with the identification of protein machinery that facilitates their function suggests that endogenous small RNAs may also possess the ability to regulate transcription. miRNAs are attractive candidates because of their well-characterized ability to regulate the expression of protein-coding genes. In addition, suggestive evidence exists that implicates a direct role for miRNAs in regulating transcription. However, prior to the work described here there had been no systematic evaluation of the potential for miRNAs to target regions outside of annotated mRNAs. The focus of this

dissertation is to test the hypothesis that miRNAs are capable of targeting gene promoters or gene termini and subsequently regulating transcription.

References For Chapter 1

1. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94 (1974).
2. Moerman, D. G. & Baillie, D. L. Genetic Organization in CAENORHABDITIS ELEGANS: Fine-Structure Analysis of the unc-22 Gene. *Genetics* **91**, 95-103 (1979).
3. Benian, G. M., L'Hernault, S. W. & Morris, M. E. Additional sequence complexity in the muscle gene, unc-22, and its encoded protein, twitchin, of Caenorhabditis elegans. *Genetics* **134**, 1097-1104 (1993).
4. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811 (1998).
5. Napoli, C., Lemieux, C. & Jorgensen, R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* **2**, 279-289 (1990).
6. van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N. & Stuitje, A. R. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* **2**, 291-299 (1990).
7. Smith, C. J. *et al.* Expression of a truncated tomato polygalacturonase gene inhibits expression of the endogenous gene in transgenic plants. *Mol Gen Genet* **224**, 477-481 (1990).
8. Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950-952 (1999).
9. Tabara, H. *et al.* The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**, 123-132 (1999).
10. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293-296 (2000).
11. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146-1150 (2001).
12. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**, 188-200 (2001).
13. Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. How cells respond to interferons. *Annu Rev Biochem* **67**, 227-264 (1998).
14. Manche, L., Green, S. R., Schmedt, C. & Mathews, M. B. Interactions between double-stranded RNA regulators and the protein kinase DAI. *Mol Cell Biol* **12**, 5238-5248 (1992).
15. Minks, M. A., West, D. K., Benvin, S. & Baglioni, C. Structural requirements of double-stranded RNA for the activation of 2',5'-oligo(A) polymerase and protein kinase of interferon-treated HeLa cells. *J Biol Chem* **254**, 10180-10183 (1979).
16. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-498 (2001).
17. Matzke, M. A., Primig, M., Trnovsky, J. & Matzke, A. J. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J* **8**, 643-649 (1989).

18. Matzke, M. A. & Matzke, A. J. Differential inactivation and methylation of a transgene in plants by two suppressor loci containing homologous sequences. *Plant Mol Biol* **16**, 821-830 (1991).
19. Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**, 567-576 (1994).
20. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* **19**, 5194-5201 (2000).
21. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833-1837 (2002).
22. Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232-2237 (2002).
23. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672-676 (2004).
24. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. *Mol Cell* **9**, 315-327 (2002).
25. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* **303**, 669-672 (2004).
26. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289-1292 (2004).
27. Ting, A. H., Schuebel, K. E., Herman, J. G. & Baylin, S. B. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat Genet* **37**, 906-910 (2005).
28. Janowski, B. A. *et al.* Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat Chem Biol* **1**, 216-222 (2005).
29. Janowski, B. A. *et al.* Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* **13**, 787-792 (2006).
30. Kim, D. H., Villeneuve, L. M., Morris, K. V. & Rossi, J. J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **13**, 793-797 (2006).
31. Chu, Y., Yue, X., Younger, S. T., Janowski, B. A. & Corey, D. R. Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. *Nucleic Acids Res* **38**, 7736-7748 (2010).
32. Han, J., Kim, D. & Morris, K. V. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci U S A* **104**, 12422-12427 (2007).
33. Yue, X. *et al.* Transcriptional regulation by small RNAs at sequences downstream from 3' gene termini. *Nat Chem Biol* **6**, 621-629 (2010).
34. Schwartz, J. C. *et al.* Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol* **15**, 842-848 (2008).
35. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
36. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351 (2001).

37. Shoemaker, D. D. *et al.* Experimental annotation of the human genome using microarray technology. *Nature* **409**, 922-927 (2001).
38. Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916-919 (2002).
39. Rinn, J. L. *et al.* The transcriptional activity of human Chromosome 22. *Genes Dev* **17**, 529-540 (2003).
40. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149-1154 (2005).
41. Kapranov, P. *et al.* Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res* **15**, 987-997 (2005).
42. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-1563 (2005).
43. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-1488 (2007).
44. Gingeras, T. R. Origin of phenotypes: genes and transcripts. *Genome Res* **17**, 682-690 (2007).
45. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
46. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007).
47. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227 (2009).
48. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* **106**, 11667-11672 (2009).
49. Hirota, K. *et al.* Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **456**, 130-134 (2008).
50. He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N. & Kinzler, K. W. The antisense transcriptomes of human cells. *Science* **322**, 1855-1857 (2008).
51. Sun, M., Hurst, L. D., Carmichael, G. G. & Chen, J. Evidence for a preferential targeting of 3'-UTRs by cis-encoded natural antisense transcripts. *Nucleic Acids Res* **33**, 5533-5543 (2005).
52. Tan-Wong, S. M., French, J. D., Proudfoot, N. J. & Brown, M. A. Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proc Natl Acad Sci U S A* **105**, 5160-5165 (2008).
53. O'Sullivan, J. M. *et al.* Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* **36**, 1014-1018 (2004).
54. Palmer, M. B., Majumder, P., Green, M. R., Wade, P. A. & Boss, J. M. A 3' enhancer controls snail expression in melanoma cells. *Cancer Res* **67**, 6113-6120 (2007).
55. Tiwari, V. K. *et al.* PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol* **6**, 2911-2927 (2008).
56. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

57. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-1465 (2002).
58. de Laat, W. & Grosveld, F. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* **11**, 447-459 (2003).
59. Patrinos, G. P. *et al.* Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev* **18**, 1495-1509 (2004).
60. Tsai, C. L., Rowntree, R. K., Cohen, D. E. & Lee, J. T. Higher order chromatin structure at the X-inactivation center via looping DNA. *Dev Biol* **319**, 416-425 (2008).
61. Tan-Wong, S. M., Wijayatilake, H. D. & Proudfoot, N. J. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev* **23**, 2610-2624 (2009).
62. Laine, J. P., Singh, B. N., Krishnamurthy, S. & Hampsey, M. A physiological role for gene loops in yeast. *Genes Dev* **23**, 2604-2609 (2009).
63. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).
64. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855-862 (1993).
65. Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901-906 (2000).
66. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86-89 (2000).
67. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853-858 (2001).
68. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**, 858-862 (2001).
69. Lee, R. C. & Ambros, V. An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**, 862-864 (2001).
70. Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**, 25-36 (2003).
71. Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83-86 (2004).
72. van Rooij, E. *et al.* Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* **316**, 575-579 (2007).
73. Williams, A. H., Liu, N., van Rooij, E. & Olson, E. N. MicroRNA control of muscle development and disease. *Curr Opin Cell Biol* **21**, 461-469 (2009).
74. Williams, A. H. *et al.* MicroRNA-206 delays ALS progression and promotes regeneration of neuromuscular synapses in mice. *Science* **326**, 1549-1554 (2009).
75. Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663-4670 (2002).
76. Zeng, Y. & Cullen, B. R. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**, 112-123 (2003).

77. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419 (2003).
78. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**, 23-34 (2001).
79. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834-838 (2001).
80. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15**, 2654-2659 (2001).
81. Nykanen, A., Haley, B. & Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309-321 (2001).
82. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563-574 (2002).
83. Schwarz, D. S., Hutvagner, G., Haley, B. & Zamore, P. D. Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. *Mol Cell* **10**, 537-548 (2002).
84. Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. & Tuschl, T. New microRNAs from mouse and human. *RNA* **9**, 175-179 (2003).
85. Lim, L. P. *et al.* The microRNAs of Caenorhabditis elegans. *Genes Dev* **17**, 991-1008 (2003).
86. Lai, E. C., Tomancak, P., Williams, R. W. & Rubin, G. M. Computational identification of Drosophila microRNA genes. *Genome Biol* **4**, R42 (2003).
87. Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**, 735-739 (2002).
88. Ghildiyal, M. *et al.* Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320**, 1077-1081 (2008).
89. Chung, W. J., Okamura, K., Martin, R. & Lai, E. C. Endogenous RNA interference provides a somatic defense against Drosophila transposons. *Curr Biol* **18**, 795-802 (2008).
90. Kawamura, Y. *et al.* Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**, 793-797 (2008).
91. Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R. & Zhu, J. K. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* **123**, 1279-1291 (2005).
92. Yang, N. & Kazazian, H. H., Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol* **13**, 763-771 (2006).
93. Katiyar-Agarwal, S. *et al.* A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A* **103**, 18002-18007 (2006).
94. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* **20**, 1709-1714 (2006).
95. Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* **20**, 1732-1743 (2006).

96. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199-202 (2006).
97. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203-207 (2006).
98. Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313**, 363-367 (2006).
99. Batista, P. J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Mol Cell* **31**, 67-78 (2008).
100. Das, P. P. *et al.* Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. *Mol Cell* **31**, 79-90 (2008).
101. Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**, 908-917 (2008).
102. Tam, O. H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534-538 (2008).
103. Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539-543 (2008).
104. Lai, E. C. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**, 363-364 (2002).
105. Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513-520 (2002).
106. Enright, A. J. *et al.* MicroRNA targets in Drosophila. *Genome Biol* **5**, R1 (2003).
107. Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of Drosophila MicroRNA targets. *PLoS Biol* **1**, E60 (2003).
108. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787-798 (2003).
109. Kiriakidou, M. *et al.* A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* **18**, 1165-1178 (2004).
110. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol* **3**, e85 (2005).
111. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105 (2007).
112. John, B. *et al.* Human MicroRNA targets. *PLoS Biol* **2**, e363 (2004).
113. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-773 (2005).
114. Olsen, P. H. & Ambros, V. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216**, 671-680 (1999).
115. Bagga, S. *et al.* Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122**, 553-563 (2005).
116. Krutzfeldt, J. *et al.* Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685-689 (2005).
117. Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75-79 (2006).

118. Rehwinkel, J. *et al.* Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in Drosophila melanogaster. *Mol Cell Biol* **26**, 2965-2975 (2006).
119. Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056-2060 (2002).
120. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-1441 (2004).
121. Wu, L., Fan, J. & Belasco, J. G. MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A* **103**, 4034-4039 (2006).
122. Behm-Ansmant, I. *et al.* mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* **20**, 1885-1898 (2006).
123. Eulalio, A. *et al.* Deadenylation is a widespread effect of miRNA regulation. *RNA* **15**, 21-32 (2009).
124. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58-63 (2008).
125. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64-71 (2008).
126. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840 (2010).
127. Hwang, H. W., Wentzel, E. A. & Mendell, J. T. A hexanucleotide element directs microRNA nuclear import. *Science* **315**, 97-100 (2007).
128. Marcon, E., Babak, T., Chua, G., Hughes, T. & Moens, P. B. miRNA and piRNA localization in the male mammalian meiotic nucleus. *Chromosome Res* **16**, 243-260 (2008).
129. Foldes-Papp, Z. *et al.* Trafficking of mature miRNA-122 into the nucleus of live liver cells. *Curr Pharm Biotechnol* **10**, 569-578 (2009).
130. Jeffries, C. D., Fried, H. M. & Perkins, D. O. Nuclear and cytoplasmic localization of neural stem cell microRNAs. *RNA* (2011).
131. Robb, G. B., Brown, K. M., Khurana, J. & Rana, T. M. Specific and potent RNAi in the nucleus of human cells. *Nat Struct Mol Biol* **12**, 133-137 (2005).
132. Rudel, S., Flatley, A., Weinmann, L., Kremmer, E. & Meister, G. A multifunctional human Argonaute2-specific monoclonal antibody. *RNA* **14**, 1244-1253 (2008).
133. Guang, S. *et al.* An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* **321**, 537-541 (2008).
134. Ohrt, T. *et al.* Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Res* **36**, 6439-6449 (2008).
135. Weinmann, L. *et al.* Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* **136**, 496-507 (2009).
136. Place, R. F., Li, L. C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* **105**, 1608-1613 (2008).
137. Kim, D. H., Saetrom, P., Snove, O., Jr. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 16230-16235 (2008).
138. Li, L. C. *et al.* Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci U S A* **103**, 17337-17342 (2006).

Chapter 2: Predicting potential miRNA target sites within gene promoters

2.1 Introduction

Synthetic small duplex RNAs complementary to gene promoters within chromosomal DNA are potent inhibitors or activators of target gene expression in mammalian cells[1-5]. Promoter-targeting small RNAs recruit members of the argonaute (AGO) protein family to RNA transcripts that originate from the target gene promoter in either the sense or antisense direction[6-9]. Recognition of the target RNA occurs in close proximity to the chromosome, resulting in transcriptional modulation of the target gene.

One remarkable feature of the synthetic promoter-targeting RNAs that we have examined is the potency and robustness of their activity when they are introduced into cells. This potency, coupled with the presence of protein machinery that facilitates their function, suggests that endogenous small RNAs may possess the ability to recognize gene promoters. If RNA could direct proteins to specific gene promoters, such RNA-mediated modulation of transcription might have evolutionary advantages relative to the development of gene-specific protein transcription factors.

Synthetic duplex RNAs that are complementary to mRNA (small interfering RNAs or siRNAs) are also potent and robust agents for modulating gene expression[10]. siRNAs are known to have endogenous analogs that regulate gene expression called microRNAs (miRNAs)[11]. miRNAs are processed inside the cell from RNA precursors that contain stem-loop structures. These stem-loop structures are processed by the double-stranded nucleases Drosha and Dicer to produce mature miRNAs.

At the initiation of this study the latest release of the miRNA repository (miRBase v12.0) contained 866 annotated human miRNAs, but this number continues to increase. Several miRNAs that recognize sequences within the 3'-untranslated regions (3'UTR) of mRNA transcripts have been characterized. Many miRNAs, however, have no known targets while some can recognize multiple mRNAs, suggesting that the determinants of miRNA interactions are complex and poorly understood[12-13].

Two reports based on computational analyses have suggested that miRNAs can modulate gene expression through promoter recognition. Dahiya and co-workers used publically available software (RegRNA) to search for potential miRNA target sites within the promoter of the E-cadherin gene[14]. They identified one potential binding site for miR-373 within the E-cadherin promoter and reported that introduction of a synthetic miR-373 mimic increased expression of the gene by 6 fold at the level of the mRNA. Rossi and co-workers searched for perfect complementarity between miRNAs and gene promoters[15]. Their analysis suggested that miR-320 targets the genomic location from which it is transcribed and showed that expression of miR-320 and the adjacent gene, POLR3D, are anti-correlated.

The above-mentioned studies either analyzed a single gene promoter or used highly stringent sequence comparison criteria. These approaches were not intended to assess broader potential for miRNAs to recognize gene promoters, warranting a more thorough evaluation of the relationship between miRNAs and promoter sequences.

A practical justification for more comprehensive studies is that validating natural gene targets of miRNAs is a complex and difficult process. The development of systematic and efficient methods for identifying promoter sequences that may be miRNA

targets is essential for prioritizing predictions and efficiently allocating experimental resources towards validating the most promising targets. Here we examine computational methods for predicting potential miRNA targets within gene promoters and demonstrate that promoters are strong candidates for miRNA regulation.

2.2 Significance

Prior to the work presented in this chapter, two reports were published describing miRNAs that are capable of recognizing sequences within gene promoters[14-15]. Both reports used computational approaches for selecting miRNAs of interest. Place *et al.* used a publicly available microRNA target prediction algorithm, RegRNA, that provides putative miRNA target sites within a supplied RNA sequence[16]. The same group had previously targeted the E-cadherin gene promoter with designed small duplex RNAs and used RegRNA to search for miRNA target sites within the same promoter[4]. This approach is effective for the analysis of a single gene promoter, but does not allow for the analysis of large datasets.

Kim *et al.* searched for miRNAs that are fully complementary to sequences within gene promoters[15]. Ten miRNAs were identified and in each case their respective predicted target sites were actually the genomic locations from which they are encoded. The authors propose a model whereby a miRNA can target its own genomic location *in cis* and regulate transcription of nearby genes. This model is restricted to a small fraction of miRNAs and does not address a broader potential for miRNAs as a class to regulate transcription.

The data presented in this chapter provide a genome-wide evaluation of putative miRNA target sites within gene promoters. This analysis allows for comparisons between traditional miRNA target sites within 3'-UTRs and potential miRNA target sites within promoter regions. Furthermore, it allows for the characterization of properties that might distinguish miRNA target sites in promoters from those in other regions of a gene. The computational tools developed here will be used in the following chapter, which is focused on the experimental validation of miRNA target sites within gene promoters.

2.3 Sequence Acquisition

To identify putative promoter-targeting miRNAs we constructed a database comprised of miRNA and gene promoter sequences from public sequence repositories. Promoter sequences were acquired from the UCSC genome browser (hg 18) and consisted of the 200 nucleotides immediately 5' to the annotated transcription start site for each gene[17-18]. We chose 200 base sequences (-200 to -1) for initial evaluations because most of the functional promoter-targeting RNAs characterized to date are designed to be complementary to this region within their respective target promoters, but larger promoter regions can also be examined. Mature miRNA sequences were obtained from miRBase (Build 12.0), which contains sequences of experimentally determined precursor and mature miRNAs[19-21].

2.4 Analysis of seed sequence matches

Synthetic promoter-targeting RNAs recognize non-coding (ncRNA) transcripts that overlap gene promoters. We used promoter DNA sequences to construct datasets

representing potential ncRNA transcripts in both the sense and antisense direction for each gene promoter as we hypothesize that endogenous small RNAs would also recognize these ncRNA transcripts. For comparison we also obtained the sequences of the 5'UTR, coding sequences (CDS), and 3'UTR for each gene (**Figure 2.1A**).

A basic requirement for target recognition by miRNAs is perfect complementarity between the target sequences and bases 2-8 of the mature miRNA sequence, called the seed sequence. We determined the number of seed matches within potential sense and antisense transcripts that overlap gene promoters and compared them to seed matches within the 3'UTR region of mRNAs (**Figure 2.1B**). We found that seed matches within promoter-overlapping transcripts occur 80% as frequently as seed matches within 3'UTRs, indicating that gene promoter sequences have the potential to be miRNA targets (**Figure 2.2**). Our analysis did detect the previously reported complementarity between miR-320 and the POLR3D promoter[15].

To evaluate the statistical significance of seed matches within gene promoter sequences we tabulated the frequency of occurrences of seed matches in 100 randomizations of each promoter sequence. We found that seed matches occur 75% as frequently within randomized as opposed to actual promoter sequences (**Figure 2.3**). The excess of observed to expected seed sequence matches within promoter sequences was similar for both putative sense and antisense transcripts. This result implies that promoter sequences are enriched for potential targets for recognition by miRNAs. Matches are equally distributed throughout the 200 base gene promoter segments surveyed, suggesting that no particular region of a gene promoter is more likely than another to contain a predicted miRNA target site (**Figure 2.4**).

2.5 Ranking matches

Our analysis identified nearly 800,000 miRNA seed matches within 27,345 gene promoter sequences (**Figure 2.3**). This large number required investigation of additional factors to prioritize target predictions. Although not necessarily a prerequisite for miRNA function, the minimum free energy (MFE) of hybridization between miRNAs and their predicted target sites have been successfully used to predict miRNA target sites within 3'UTRs[22]. We reasoned that MFE values may also be useful for prioritizing miRNA target predictions within gene promoters.

The MFE values were calculated for miRNA hybridization to predicted target sites (based on seed sequence matches, hereafter simply referred to as predictions) within putative promoter-overlapping transcripts and within 10 randomizations of promoter sequences. We found that predictions with lower MFE values occurred more frequently in actual promoter sequences than in randomized sequences (**Figure 2.5**). The difference between the distributions of MFE values demonstrates that predictions with low MFE values occur more often than would be expected at random, implying that these predictions are more likely to be biologically significant and that MFE values will be useful criteria for prioritizing target predictions.

During the course of the MFE analysis we identified several miRNA target predictions within gene promoters that had notably low MFE values. These observations prompted us to compare the MFE values for target predictions within gene promoters to target predictions within 3'UTRs (**Figure 2.6**). We calculated the mean MFE value for all predictions within gene promoters to be -24.27 kcal/mol and -24.32 kcal/mol for putative sense and antisense promoter-overlapping transcripts, respectively. The mean

MFE for all predictions within 3'UTRs was -20.57 kcal/mol, more than 3.5 kcal/mol higher than predictions within promoters. The difference in mean MFE values suggests that, on average, miRNA recognition of sequences at gene promoters would be more energetically favorable than recognition of 3'UTR sequences.

To further evaluate the differences between target predictions within gene promoters and 3'UTRs, we examined the distribution of MFE values for all predictions within the different sequence datasets. As previously indicated by the mean MFE values, roughly 50% of target predictions within gene promoters had MFE values below -24.3 kcal/mol. Interestingly, only 22% of predictions within 3'UTRs had MFE values below -24.3 kcal/mol (**Figure 2.6**). The difference in MFE value distributions demonstrates that gene promoters are enriched relative to 3'UTRs for predicted target sites with low free energies of hybridization and may actually represent more favorable miRNA targets than 3'UTRs.

Another criterion used in miRNA target prediction is sequence complementarity. Sequence complementarity alone has been used successfully to predict miRNA target sites within 3'UTRs[23]. We used the Needleman-Wunsch algorithm to evaluate the degree of sequence complementarity between miRNAs and predicted target sites within gene promoters (**Figure 2.1B**)[24]. We identified over 200 individual miRNAs with near perfect complementarity to their predicted target sites within gene promoters. A selected subset of these predictions is listed in **Figure 2.7**. The high degree of complementarity between miRNAs and gene promoters further demonstrates that gene promoters are promising candidates for miRNA targets.

2.6 Conclusion

Strong evidence that gene expression can be modulated using synthetic duplex RNAs that are complementary to gene promoters suggests that natural gene regulation may include recognition of gene promoters by miRNAs. Such recognition would have evolutionary advantages, given the large difference between protein transcription factors and miRNAs in their efficiency of generating new selectivity for gene promoters through mutation.

Here we describe a computational algorithm that can be used to identify promising miRNA target sites within gene promoters. We identify many seed sequence matches within promoters and demonstrate that they are almost as common as those within 3'UTRs. We also identify many miRNA/promoter pairs that have unusually strong complementarity. These results can be used to rank order miRNA/promoter pairs for the demanding studies necessary to validate whether the potential for these interactions is biologically significant.

Figure 2.1. Computational approach for identifying miRNAs that target gene promoters. (A) Diagram of sequences analyzed for miRNA target sites. (B) Schematic of algorithm used to predict miRNA targets within gene promoters.

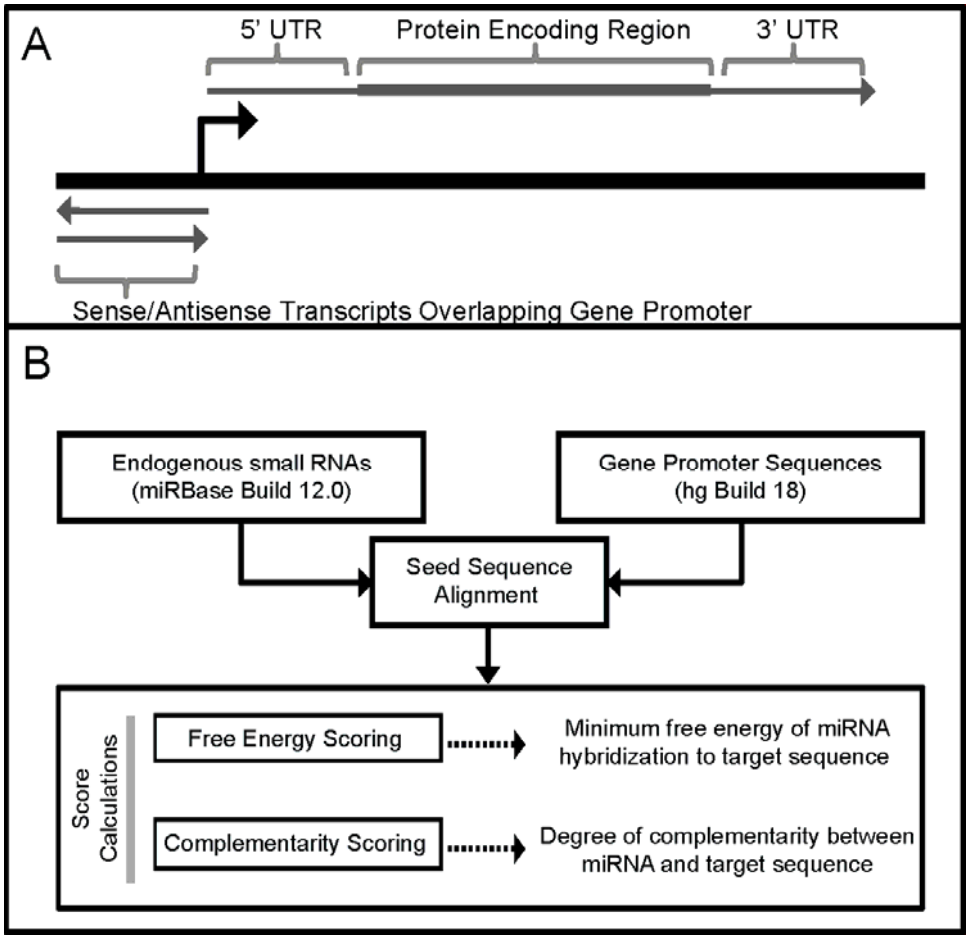


Figure 2.2. The frequency of seed sequence matches in promoter regions, 5'UTRs, coding regions, and 3'UTRs.

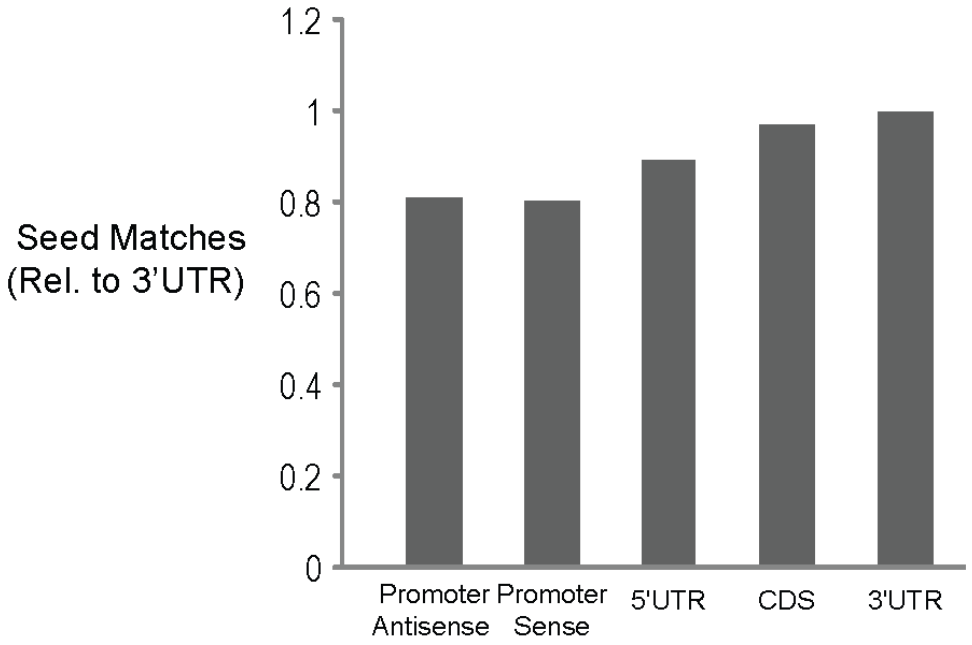


Figure 2.3. Comparison of seed matches within promoter-overlapping transcripts vs. randomized promoter sequences. ($P < .01$)

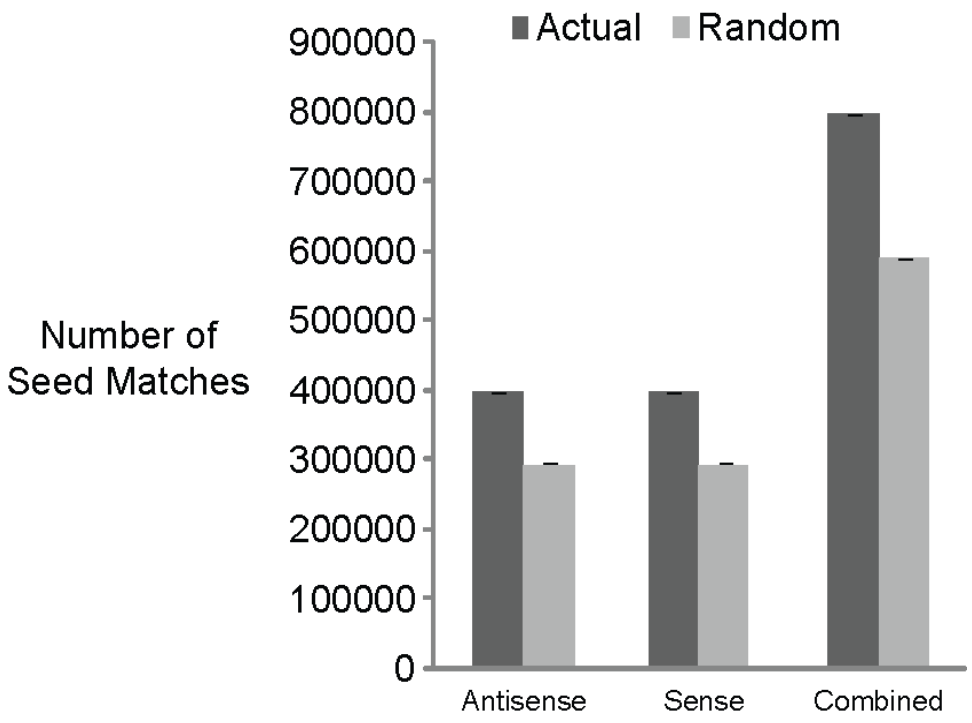


Figure 2.4. Distribution of seed match locations within sense and antisense transcripts that overlap gene promoters from -1 to -200 relative to the +1 transcription start site.

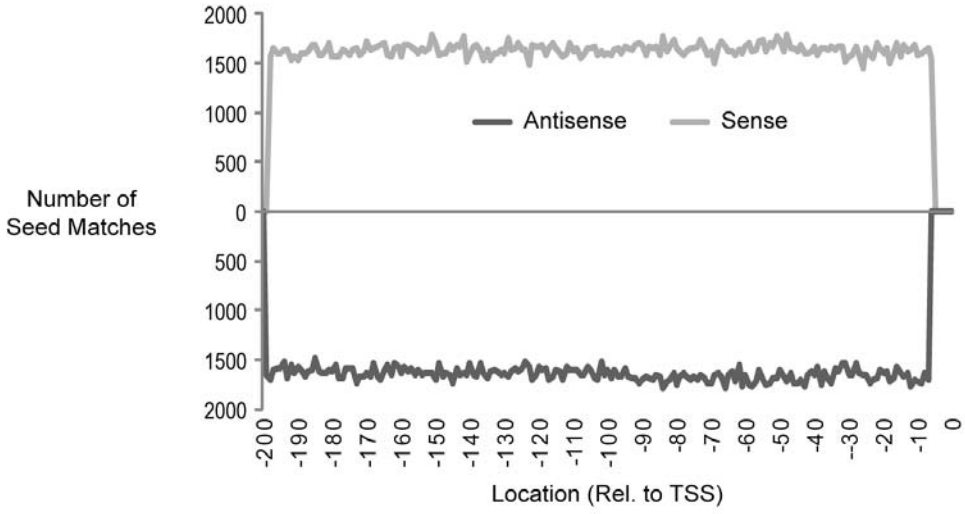


Figure 2.5. Distribution of MFE values for predictions within promoter-overlapping transcripts as compared to randomized promoter sequences.

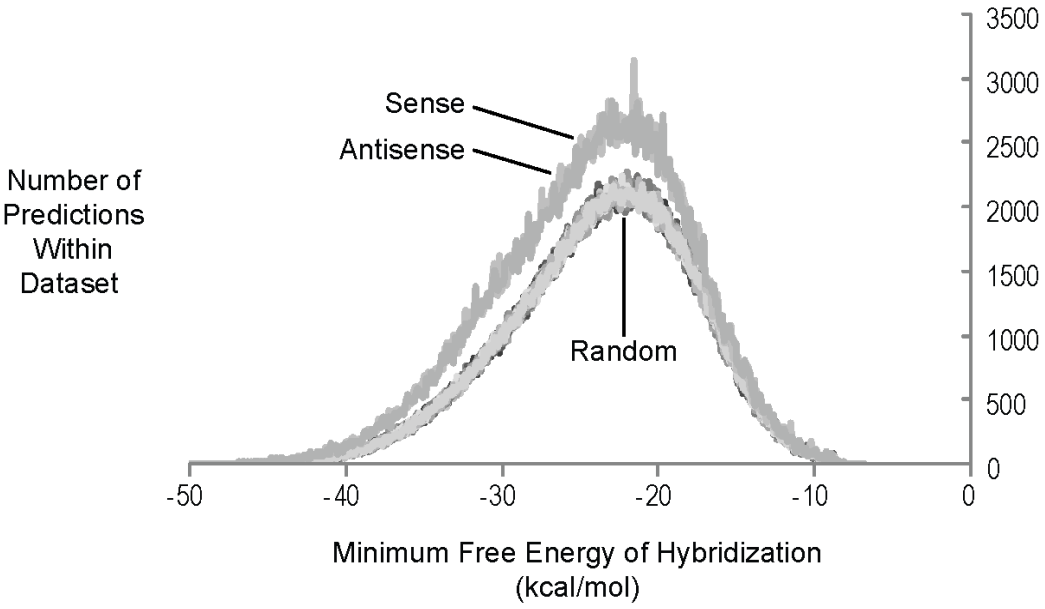


Figure 2.6. Distribution of MFE values for predictions within promoter-overlapping transcripts as compared to 3'UTRs.

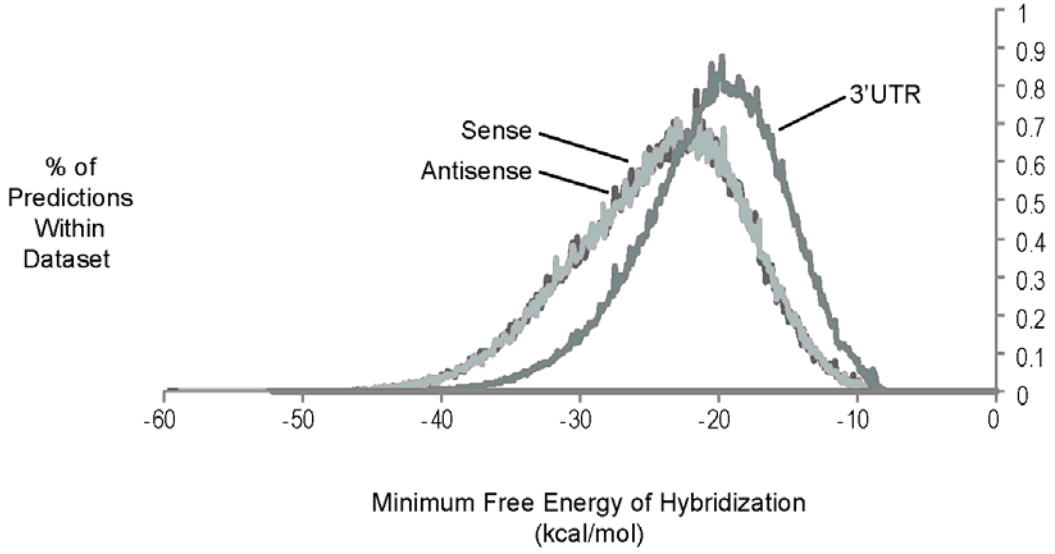
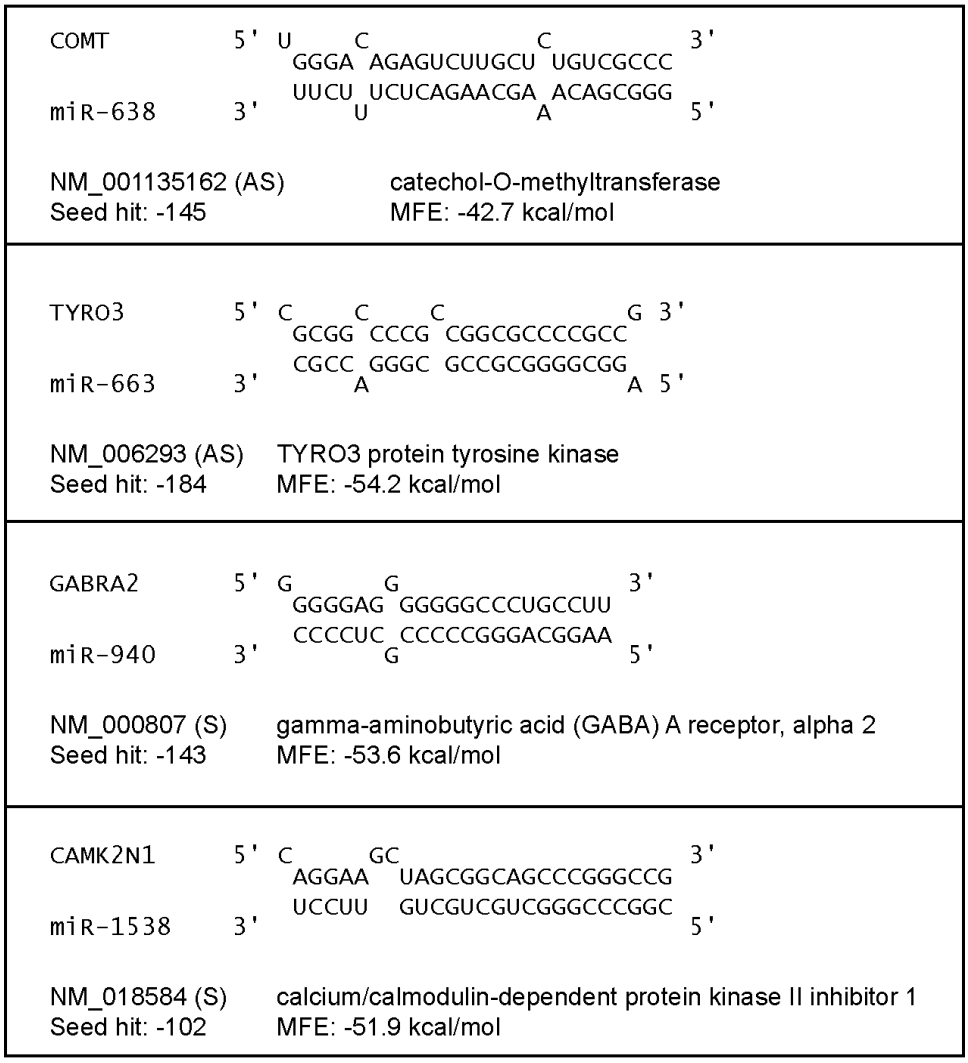


Figure 2.7. Examples of predicted miRNA targets within sequences of promoter-overlapping transcripts.



References For Chapter 2

1. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289-1292 (2004).
2. Ting, A. H., Schuebel, K. E., Herman, J. G. & Baylin, S. B. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat Genet* **37**, 906-910 (2005).
3. Janowski, B. A. *et al.* Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat Chem Biol* **1**, 216-222 (2005).
4. Li, L. C. *et al.* Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci U S A* **103**, 17337-17342 (2006).
5. Janowski, B. A. *et al.* Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol* **3**, 166-173 (2007).
6. Janowski, B. A. *et al.* Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* **13**, 787-792 (2006).
7. Kim, D. H., Villeneuve, L. M., Morris, K. V. & Rossi, J. J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **13**, 793-797 (2006).
8. Han, J., Kim, D. & Morris, K. V. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci U S A* **104**, 12422-12427 (2007).
9. Schwartz, J. C. *et al.* Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol* **15**, 842-848 (2008).
10. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811 (1998).
11. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853-858 (2001).
12. John, B. *et al.* Human MicroRNA targets. *PLoS Biol* **2**, e363 (2004).
13. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).
14. Place, R. F., Li, L. C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* **105**, 1608-1613 (2008).
15. Kim, D. H., Saetrom, P., Snove, O., Jr. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 16230-16235 (2008).
16. Huang, H. Y., Chien, C. H., Jen, K. H. & Huang, H. D. RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res* **34**, W429-434 (2006).
17. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
18. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
19. Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res* **32**, D109-111 (2004).

20. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-144 (2006).
21. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**, D154-158 (2008).
22. Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of Drosophila MicroRNA targets. *PLoS Biol* **1**, E60 (2003).
23. Lai, E. C. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**, 363-364 (2002).
24. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443-453 (1970).

Chapter 3: Transcriptional Gene Silencing in Mammalian Cells by miRNAs That Target Gene Promoters

3.1 Introduction

Small duplex RNAs complementary to gene promoters are potent silencers and activators of target gene expression in mammalian cells[1-12]. We and others have demonstrated that these small RNAs sequence-specifically recognize non-coding RNA (ncRNA) transcripts that overlap gene promoters[6,12-15]. Recognition of the target ncRNA occurs in close proximity to the chromosome and leads to recruitment of Argonaute (AGO) proteins to the ncRNA and to the target gene promoter[13-14,16-18]. Modulation of target gene expression occurs at the level of transcription and has been observed in both the presence and absence of DNA methylation[1-7].

Potent and robust transcriptional modulation suggests the existence of an endogenous mechanism that facilitates recognition of gene promoters by small RNAs. AGO proteins implicated in the mechanism of promoter-targeting RNAs are conserved across eukaryotes[19]. AGO proteins have been shown to use endogenous small RNAs to direct chromatin modification and regulate transcription in yeast and plants[20-22]. In humans, AGO proteins are well known for their role in microRNA (miRNA) regulation. miRNAs are an endogenous class of short regulatory RNAs ranging in length from 19-30 bases. They are generally understood to repress gene expression post-transcriptionally through recognition of sequences within 3'-UTRs of mRNA transcripts[23-24]. miRNAs have been reported to regulate transcription in human cells, but much remains to be learned about their mechanism of action[25-26].

We previously performed a genome-wide evaluation of gene promoters and found a significant enrichment of putative miRNA target sites within promoter regions[27]. Using the human progesterone receptor (PR) as a model gene we have now identified multiple miRNAs that are predicted to target the PR promoter and inhibit PR expression. We further demonstrate that a ncRNA transcribed from the PR promoter serves as the molecular target of these miRNAs. Our results indicate that miRNA recognition of gene promoters may be a general mechanism for gene regulation.

3.2 Significance

Prior to the work presented in this chapter, there had been two reports published that described miRNAs that are capable of targeting gene promoters and modulating transcription in human cells[25-26]. Place *et al.* reported that a miRNA mimic with incomplete complementary to the E-cadherin promoter activated expression of the gene. The authors demonstrated that gene activation was associated with increased occupancy of RNA Pol II on the E-cadherin promoter, but provided no direct evidence for an interaction between the miRNA and the targeted promoter. Furthermore, inhibition of the endogenous miRNA had no effect on E-cadherin expression.

Kim *et al.* reported that a miRNA can target its own genomic location and silence transcription of an adjacent gene[26]. The authors showed that expression of one specific miRNA is anti-correlated to the expression of a gene that is encoded immediately downstream of the miRNA within the genome. In addition, one experiment showed that introduction of a miRNA mimic resulted in an increased occupancy of AGO1 and H3K27me3 on the targeted promoter. In yeast, endogenous siRNAs target their own

genomic location and recognition results in the formation of constitutive heterochromatin[21,28]. However, the gene studied in the Kim *et al.* report is located within a euchromatic region of the genome and the biological rationale for such a mechanism is unclear.

The data presented in this chapter provide a systematic evaluation of the ability of miRNAs to target sequences within gene promoters. As opposed to the previous studies that focused on single miRNAs, we screen multiple miRNAs to better understand their ability to recognize promoter sequences[25-26]. We identify several miRNAs that are capable of targeting a common gene promoter. In addition, we characterize one miRNA that is capable of targeting multiple promoters. Collectively, our observations establish a generality for miRNA-induced transcriptional gene silencing that was previously unappreciated.

3.3 miRNAs Complementary to the PR Promoter Inhibit PR Expression

We have previously designed small duplex RNAs complementary to the progesterone receptor (PR) promoter and demonstrated that these RNAs silence transcription of the PR gene[4,16]. We subsequently showed that the molecular targets of these small RNAs are ncRNA transcripts produced from the PR promoter[14]. The PR gene has two major isoforms termed PR-B and PR-A that differ in their transcription start sites (TSS), PR-B being the most upstream. One of the ncRNAs that we characterized is a 2170-base transcript that initiates 1431 bases downstream of the PR-B TSS and is transcribed in the antisense direction through the PR promoter.

In prior studies we used designed synthetic small duplex RNAs that were fully complementary to the PR promoter. The differences between fully complementary designed RNAs and miRNAs are substantial. miRNAs, which are encoded within the genome, do not require complete complementarity to achieve target recognition. This makes it difficult to accurately predict sequences miRNAs will efficiently target. To begin to experimentally determine if miRNAs are capable of recognizing sequences within gene promoters, we obtained sequences for all known human miRNAs from miRBase, the public repository for miRNAs[29]. Using an algorithm we developed to identify potential miRNA target sites, we scanned the segment of the ncRNA that overlaps the PR-B promoter for potential miRNA target sites[27]. We identified 72 potential miRNA target sites within the analyzed region of the ncRNA transcript. We selected ten miRNAs for experimental validation based on complementarity to their respective target sequences within the PR promoter (**Figure 3.1**, **Figure 3.2**). We designed miRNA mimics that consisted of the miRNA sequence and a fully complementary RNA carrier strand. The miRNA mimics were transfected into T47D breast cancer cells and PR protein expression was monitored by Western blot. PRC1 (an siRNA targeting PR mRNA) and PR9 (a duplex RNA previously shown to target the PR promoter and inhibit PR expression) were used as positive controls. As a negative control we used a mismatched duplex RNA (MM) that does not affect PR expression.

Several miRNA mimics inhibited PR protein expression (**Figure 3.3A**). Quantification of replicate experiments revealed that inhibition by miR-372, miR-373, miR-520c-3p, and miR-423-5p was statistically significant (**Figure 3.3B**). These results indicate that miRNAs complementary to the PR promoter are capable of inhibiting PR

expression. In striking contrast to our prior experience with designed synthetic small RNAs, where introduction of more than one mismatched base between the silencing RNA and its target completely abolished activity, all of the miRNA mimics that inhibited PR expression were extensively mismatched when aligned to the target ncRNA (**Figure 3.2**). For example, miR-423-5p possesses seven mismatched bases relative to its ncRNA target, including a three base bulge adjacent to the seed sequence. miR-520c-3p also contains 7 mismatched bases relative to the target ncRNA, but with a five base bulge immediately following the seed sequence.

The target sites for the inhibitory miRNAs also differ significantly from the sites we have previously targeted with fully complementary small duplex RNAs. In our prior studies, any duplexes that targeted more than 26 bases upstream of the PR-B TSS were inactive[4]. However, three of the inhibitory miRNA mimics we identified target more than 500 bases upstream of the TSS. This suggests that small RNAs can target sequences much further upstream of the PR TSS than has previously been appreciated. The most potent mimic, miR-423-5p, targets the region spanning from 59 bases upstream of the PR-B TSS to 37 bases upstream of the TSS. miR-423-5p inhibited PR expression in a dose dependent manner with an IC_{50} of 7.2 nM (**Figure 3.4**) and was selected as the focus for subsequent experiments.

3.4 miR-423-5p Inhibits Transcription of the PR Gene

Potent inhibition of PR expression by miR-423-5p prompted us to investigate the mechanism of silencing. Using quantitative RT-PCR (RT-qPCR) we found that miR-423-5p reduced PR mRNA levels by more than 70% (**Figure 3.5A**). To evaluate

whether decreased mRNA levels are due to reduced transcription we monitored expression of pre-spliced mRNA, also termed heteronuclear RNA (hnRNA), produced from the PR locus. miR-423-5p reduced hnRNA levels by nearly 70% (**Figure 3.5B**). We also quantified RNA Polymerase II (Pol II) occupancy at the PR promoter using chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR). Treatment with miR-423-5p resulted in a greater than 60% decrease in Pol II on the PR promoter as compared to a negative control duplex (**Figure 3.6A**). Transcriptional silencing by small RNAs has been associated with induction of repressive histone marks, including dimethylation of histone H3 lysine 9 (H3K9me2)[17]. Using ChIP-qPCR we observed a 2.3-fold increase in H3K9me2 following treatment with miR-423-5p (**Figure 3.6B**). Taken together, these results suggest that miR-423-5p represses transcription of the PR gene.

To ensure that inhibition of PR by miR-423-5p was not an artifact of our miRNA mimic design, we designed a pre-miR-423 mimic. The pre-miR-423 mimic retained the structure of endogenous miR-423 after DICER processing (an imperfect duplex RNA consisting of both the miR-423-5p and miR-423-3p sequences) (**Figure 3.7A**). Inhibition of PR expression by the pre-miR-423 mimic was indistinguishable from the miR-423-5p mimic (**Figure 3.7B**). To further verify that both the miR-423-5p and pre-miR-423 mimics were being processed appropriately we monitored miRNA levels using RT-qPCR. Addition of pre-miR-423 increased detection of both miR-423-5p and miR-423-3p by several orders of magnitude (**Figure 3.8A**) while the miR-423-5p mimic only increased detection of miR-423-5p (**Figure 3.8B**). To confirm that the miRNA mimics were entering the nucleus we monitored miR-423 levels in nuclear fractions of

transfected cells. We observed a 200-fold increase in nuclear miR-423-5p levels following treatment with the miR-423-5p mimic (**Figure 3.8C**). Similar results were obtained with the pre-miR-423 mimic (**Figure 3.8D**). These data suggest that our miRNA and pre-miRNA mimics work as designed and are capable of entering the nucleus. However, lipid-based transfection methods do result in compartmentalization of exogenously added nucleic acid and it is unclear how much of the detected miRNA is accessible to cellular machinery and functional.

In parallel to miR-423-5p we tested additional inhibitory miRNA mimics miR-372, miR-373, and miR-520c-3p for their effects on expression of PR mRNA. All three mimics reduced PR mRNA expression by more than 70% (**Figure 3.9A**). The mimics also reduced levels of PR hnRNA expression (**Figure 3.9B**). The ability of multiple promoter-targeting miRNA mimics to inhibit PR expression suggests that recognition of gene promoter sequences may be a general property of miRNAs.

3.5 Sequence Analysis of the miR-423-5p Target Site Within the PR Promoter

miR-423-5p has a high degree of complementarity to an antisense transcript that overlaps the PR gene promoter (**Figure 3.10A**). The region of complementarity spans from 59 bases upstream of the PR TSS through 37 bases upstream of the TSS. PhastCons analysis of the multiz alignment of 44 vertebrate species revealed that this region of the PR promoter is highly conserved, implying that the target sequence is of biological importance (**Figure 3.10B**).

A major determinant of miRNA target recognition is complete complementarity between the miRNA seed sequence (bases 2-8 of the mature miRNA) and the target RNA

sequence. One component of the algorithm we used to identify candidate miRNAs was the requirement for seed sequence complementarity with the ncRNA overlapping the PR promoter. Interestingly, we observed that the entire miR-423-5p seed sequence target site is highly conserved within the PR promoter. In addition to seed sequence complementarity, we noted two additional regions of complementarity between miR-423-5p and the ncRNA (**Figure 3.10A**).

To determine the importance of each segment of complementarity on gene silencing, we designed mutant miRNA mimics that contained three-base changes in either the seed sequence (Seed-MM), the middle region of complementarity (Mid-MM), or the terminal region of complementarity (End-MM) (**Figure 3.11A**). We used RT-qPCR to measure expression of PR mRNA following treatment with each mutant miRNA mimic. Mutating the seed sequence abolished silencing activity while mutating the middle or end regions of complementarity had little or no effect on silencing, respectively (**Figure 3.11B**). Identical results were obtained when evaluating the effects of these mutant mimics on PR protein expression (**Figure 3.11C and D**). These results are consistent with known targeting properties of miRNAs and demonstrate that the seed sequence of miR-423-5p is a major determinant of target recognition.

It has been reported that specific cellular RNAs are capable of sequestering DNA-binding transcription factors[30]. To test whether miR-423-5p sequesters specific protein factors that normally recognize the conserved DNA sequence within the PR promoter, we designed a small duplex DNA corresponding to the miR-423-5p sequence. The miR-423-5p DNA duplex did not affect PR mRNA expression, indicating that

transcriptional silencing is specific to small RNAs and most likely not a result of sequestering putative DNA binding proteins (**Figure 3.12**).

3.6 Mechanism of miRNA-Induced Transcriptional Silencing

Small RNA-induced transcriptional silencing has been reported in both the presence and absence of DNA methylation[1-7]. There is a CpG island near the PR-B TSS that has the potential to become methylated. We examined the methylation state of this CpG island using bisulfite treatment followed by methylation specific PCR (MSP). As a positive control we used genomic DNA isolated from MDA-MB-231 breast cancer cells where the PR promoter is methylated. Treatment with miR-423-5p did not induce DNA methylation at the PR promoter (**Figure 3.13A**). Bisulfite sequencing confirmed the lack of DNA methylation throughout the PR promoter (**Figure 3.13B**). Consistent with the absence of DNA methylation, silencing of PR protein expression by miR-423-5p was transient, lasting approximately 7-8 days before beginning to return to normal levels (**Figure 3.14**). The mechanism of transcriptional silencing by miR-423-5p at the PR promoter does not appear to involve induction of permanent epigenetic changes.

We have previously demonstrated that synthetic small duplex RNAs targeting the PR promoter recruit AGO2 to the antisense ncRNA transcript overlapping the PR promoter[16,18]. Likewise, other groups have reported the involvement of AGO1 in small RNA-induced transcriptional gene silencing[17]. We used RNA immunoprecipitation (RIP) to evaluate both AGO1 and AGO2 occupancy on the ncRNA following treatment with miR-423-5p. RIP revealed no association between AGO1 and the ncRNA under any treatment condition (**Figure 3.15A**). Conversely, treatment with miR-423-5p resulted in

association of AGO2 with the ncRNA transcript (**Figure 3.15B**). We used RT-qPCR to determine if recruitment of AGO2 affected expression of the ncRNA overlapping the PR promoter. Addition of miR-423-5p resulted in a 30% decrease in expression of the ncRNA suggesting that the miRNA moderately destabilized the target ncRNA (**Figure 3.16**). These results indicate that the ncRNA that overlaps the PR promoter is a target for miR-423-5p.

To further evaluate the necessity of AGO proteins for inhibiting PR by miR-423-5p we used siRNAs to knock down expression of either AGO1 or AGO2 prior to miRNA mimic addition (**Figure 3.17A and C**). Depletion of AGO1 had no effect on PR expression or the inhibition of PR by miR-423-5p (**Figure 3.17B**). Knockdown of AGO2 caused a two-fold increase in PR expression (**Figure 3.17D**). While we routinely observe a 75% decrease in PR expression by miR-423-5p, PR expression was only reduced by 35% in an AGO2 depleted background. AGO2 expression was required for inhibition by both the miR-423-5p mimic and the pre-miR-423 mimic suggesting that our mimic design did not bias the AGO isoform used for inhibition of PR expression (**Figure 3.17B and D**).

3.7 Effects of miR-423-5p inhibition on PR expression

Our approach in this investigation has been to screen for miRNAs that are capable of silencing transcription using miRNA mimics as a proof-of-principle for miRNA-induced transcriptional silencing and subsequently characterize the mechanism of action. An alternative approach would be to inhibit endogenously expressed miRNAs and evaluate the effects on expression of genes that contain predicted target sites within their

promoters. This is more complicated because in addition to the computational target predictions, knowledge about endogenous miRNA expression is required and the endogenous control pathway must be present in the cell types used for the experiment. Furthermore, the tools available for inhibiting miRNAs are not as well characterized as those for mimicking their function. However, miR-423-5p is expressed at detectable levels in T47D cells so we evaluated whether the endogenous miRNA was regulating PR expression.

To inhibit the function of endogenous miR-423-5p we designed an antisense 2'-O-methyl RNA that was fully complementary to the miRNA (anti-miR). As controls we also designed 2'-O-methyl RNAs consisting of either mismatched (MM) or scrambled (SCR) derivatives of the anti-miR sequence. RT-qPCR confirmed that addition of the anti-miR significantly reduced detection of miR-423-5p (**Figure 3.18A**). Inhibition of miR-423-5p did not affect PR mRNA expression (**Figure 3.18B**). This result is not particularly surprising as T47D cells express high basal levels of the PR gene and it's unlikely that the gene is under miRNA-mediated repression in this cell line.

The PR gene is expressed at low basal levels in MCF7 breast cancer cells and we reasoned that it might be a better cell line for studying the effects of endogenous miRNAs on PR expression. Using RT-qPCR we were able to detect miR-423-5p expression in MCF7 cells and used the anti-miR to inhibit the function of the endogenous miRNA (**Figure 3.18C**). Addition of the anti-miR had no effect on PR mRNA expression (**Figure 3.18D**). These data suggest that while miR-423-5p is capable of silencing transcription of the PR gene, the endogenous miRNA is not performing this function in

the cell lines we tested and that a physiological role for miR-423-5p in PR regulation, if any, will be found in other cell types or under other physiologic conditions.

3.8 Generality of miRNA-Induced Transcriptional Silencing

The potency of miR-423-5p-induced silencing of the PR gene in T47D cells led us to test the effect of miR-423-5p on PR expression in MCF7 cells. MCF7 cells express PR at low basal levels as compared to T47D cells[9]. We transfected miR-423-5p into MCF7 cells and monitored PR expression. Western blot revealed that miR-423-5p inhibited PR protein expression (**Figure 3.19A**). Using RT-qPCR we observed a greater than 60% reduction in PR mRNA expression (**Figure 3.19B**). In addition, miR-423-5p resulted in a greater than 70% reduction in PR hnRNA levels (**Figure 3.19C**). These data demonstrate that miRNA-induced transcriptional silencing of PR is readily observed in two human cell lines with differing basal levels of PR expression.

To further establish the generality of miRNA-induced transcriptional silencing we searched for additional genes with miR-423-5p target sites within their promoters. Sequence analysis revealed a miR-423-5p target site within a conserved region of the immunoglobulin superfamily member 1 (IGSF1) gene promoter (**Figure 3.20A**). miR-423-5p is complementary to a putative promoter-overlapping ncRNA transcribed in the antisense direction relative to the IGSF1 gene at a region spanning from 36 bases to 8 bases upstream of the IGSF1 TSS. IGSF1 is expressed at very low levels in T47D cells, however RT-PCR readily detected expression of IGSF1 mRNA in MCF7 cells (**Figure 3.20B**). We also detected expression of an RNA species overlapping the IGSF1 promoter that included the miR-423-5p target sequence (**Figure 3.20B**).

We transfected miR-423-5p into MCF7 cells and monitored IGSF1 expression using RT-qPCR. We observed a 60% reduction in IGSF1 mRNA expression after treatment with miR-423-5p (**Figure 3.20C**). In addition, miR-423-5p caused a greater than 70% reduction in IGSF1 hnRNA levels (**Figure 3.20D**). We next tested the effects of our previously designed mutant miR-423-5p mimics on IGSF1 mRNA expression (**Figure 3.21A**). Just as we observed with PR, silencing of IGSF1 by miR-423-5p required seed sequence complementarity to the predicted target RNA (**Figure 3.21B**). Also consistent with our previous results, expression of IGSF1 was significantly inhibited by pre-miR-423 (**Figure 3.22**). These results underscore the potential of single miRNAs to target multiple gene promoters and regulate transcription.

3.9 Conclusion

miRNAs are powerful regulators of gene expression that function by recognizing complementary RNA sequences. The vast majority of studies on miRNAs have focused on their ability to target sequences within mRNA. Here we show that miRNAs can also recognize ncRNAs overlapping gene promoters and regulate transcription. We have identified multiple promoter-targeting miRNAs that inhibit PR expression, suggesting that transcriptional regulation may be a general property of miRNAs. We have also characterized one miRNA in detail and demonstrated that it can target additional gene promoters, further supporting the possibility that recognition of gene promoters by miRNAs may be a general mechanism of gene regulation.

Our findings differ from our previous observations of transcriptional silencing by designed small RNAs. For example, our earlier reports focused on the region spanning

just 50 bases upstream of the PR-B TSS and no duplexes that targeted more than 26 bases upstream of the TSS were active. However, three of the inhibitory miRNAs we have identified here target more than 500 bases upstream of the TSS (**Figure 3.1**). This report is also the first time we have used duplex RNAs with two imperfectly complementary RNA strands and observed gene silencing (**Figure 3.7**). This result demonstrates that small duplex RNAs do not require full complementarity in order to be processed into the RNAi machinery.

Perhaps the most striking observation in our current study is the degree of imperfect complementarity between the inhibitory miRNAs and their target ncRNA. In our prior studies we designed small duplex RNAs that were fully complementary to the PR promoter and the activity of these silencing RNAs was completely abolished when duplexes contained more than one mismatched base from their target sequence[4]. In contrast, all of the inhibitory miRNAs we identified were extensively mismatched relative to their target sequence (**Figure 3.2**).

Small RNAs with imperfect complementarity to their targets face challenges that differ substantially from those that have full complementarity. miRNAs with imperfect complementarity likely have greatly reduced affinity for their target sequences (as compared to fully complementary RNAs) which would increase their difficulty achieving potent and selective binding to targets. To overcome this hurdle, at least in the case of traditional miRNA targeting, 3'-UTRs containing miRNA binding sites with poor complementarity often have multiple miRNA binding sites within the same UTR to promote cooperativity[31]. The miRNAs identified in our study by contrast, are potent inhibitors of transcription despite having several mismatched bases with their target

sequences and only one target site within the ncRNA. This result implies that the rules for miRNA target recognition within promoter regions may differ with respect to target site multiplicity and cooperativity.

There have been previous reports of miRNAs targeting gene promoters in human cells. The first report of a promoter-targeting miRNA, from Place *et al*., showed that introduction of exogenous miR-373 induced expression of two genes that contain promoter regions with complementarity to the miRNA[25]. Subsequently, Kim *et al*. reported that miR-320 targets its own genomic location *in cis* and silences transcription of an adjacent gene, POLR3D through an AGO1-dependent mechanism[26]. Our data differ significantly from these previous results. Because miR-320 targets its own genomic location, it inherently has full complementarity to its target sequence. We demonstrate that miRNAs do not require full complementarity to recognize their targets within gene promoters and silence transcription. We also show that gene silencing can be achieved by miRNAs that target gene promoters *in trans* and that a single miRNA can silence transcription of multiple genes. Finally, whereas miR-320 was reported to recruit AGO1 to its target, we find that miRNAs with incomplete complementarity to their targets require AGO2, suggesting the potential for different silencing mechanisms.

Our work provides insights into the mechanism of miRNA-induced transcriptional silencing that were not reported in previous studies. We show that miRNAs can recruit AGO2 to ncRNA transcripts that overlap their target gene promoter, implicating recognition of ncRNA transcripts in the mechanism of promoter-targeting miRNAs (**Figure 3.15B**). In our model system, recognition of the target ncRNA by the miRNA results in moderate reduction in expression of the ncRNA (**Figure 3.16**). We do

not believe that the reduction in ncRNA expression is a major factor in the silencing mechanism. In support of this conclusion, we have previously shown that depletion of the target ncRNA using an RNase H based approach (which is independent of the RNAi machinery) does not affect PR expression[14]. Instead, we believe that AGO2 is recruiting other protein factors to the PR promoter that affect transcription of the PR gene. For example, we observed an increase in H3K9 dimethylation following treatment with miR-423-5p suggesting that some of these factors may be chromatin-modifying enzymes (**Figure 3.6B**).

Transcriptome studies have revealed that over 70% of gene promoters are overlapped by ncRNA transcripts[32]. One function of these ncRNAs may be to serve as targets for miRNAs, which would be consistent with our previous observation that gene promoter sequences are enriched with potential miRNA target sites[27]. Transcriptome studies also suggest that over 90% of the human genome is transcribed into RNA, raising the possibility that additional non-coding regions of the genome could serve as targets for miRNAs[32]. In support of this hypothesis, we have recently demonstrated that synthetic small duplex RNAs can target regions beyond the 3'-terminus of protein coding genes and regulate transcription of the upstream gene[33]. These small RNAs recognize ncRNA transcripts that overlap gene termini in a manner similar to promoter-targeting small RNAs and subsequently regulate transcription through long-range chromosome interactions.

miRNAs have been understood to regulate gene expression at the post-transcriptional level through recognition of 3'-UTRs within mRNA transcripts. Our study extends miRNA function to recognition of ncRNA transcripts that overlap gene

promoters. Sequence specific recognition of gene promoters by miRNAs may complement protein transcription factors. In addition, the ability of small RNAs to rapidly evolve specificity for new sequences would have evolutionary advantages.

Figure 3.1. Schematic of selected miRNAs complementary to the PR promoter.

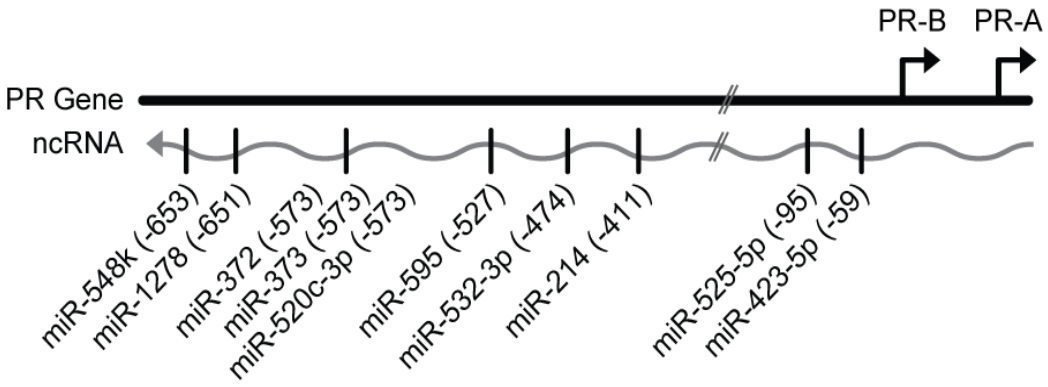


Figure 3.2. Alignments of selected miRNAs with the target ncRNA overlapping the PR promoter. Seed sequences shown in red.

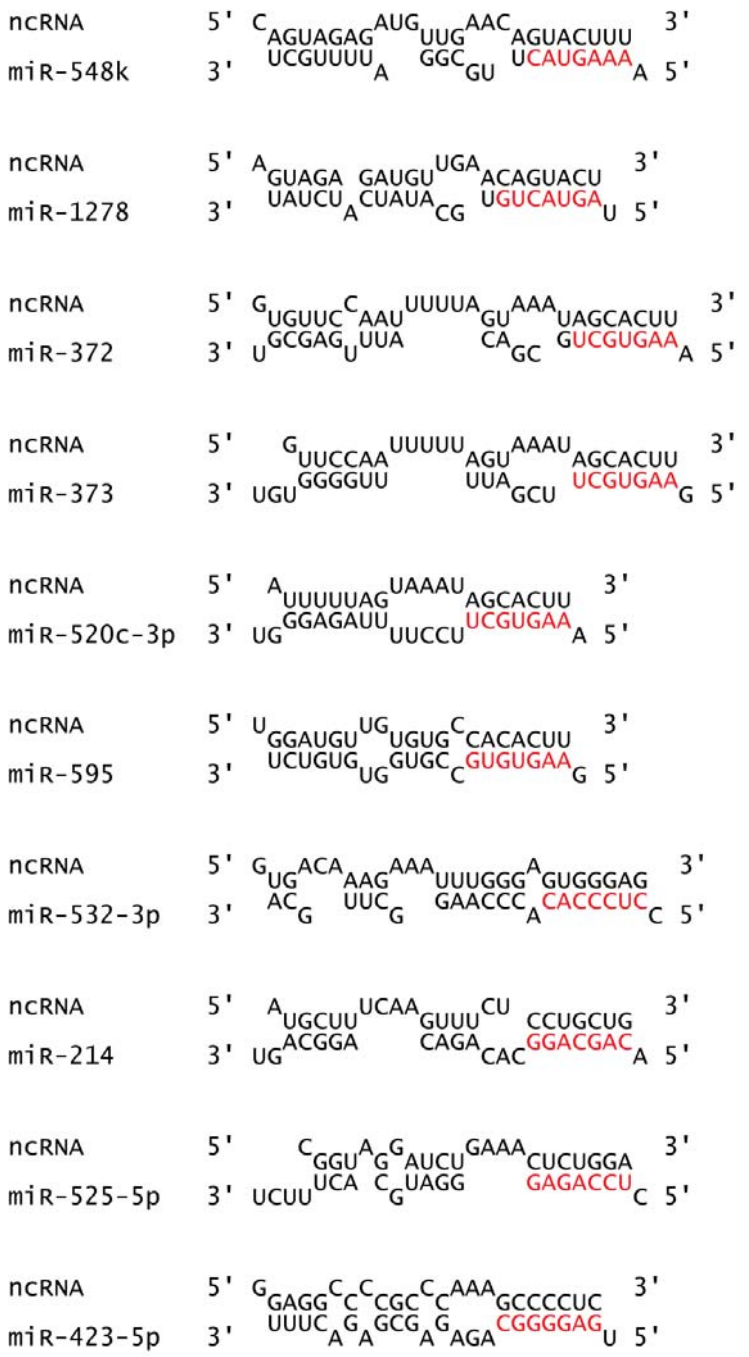


Figure 3.3. miRNAs complementary to the PR promoter inhibit PR protein expression. (A) Western analysis showing inhibition of PR protein expression by miRNA mimics. (B) Quantification of three independent experiments measuring reduction of PR protein expression by miRNA mimics. miRNA mimics were added to cells at 50 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

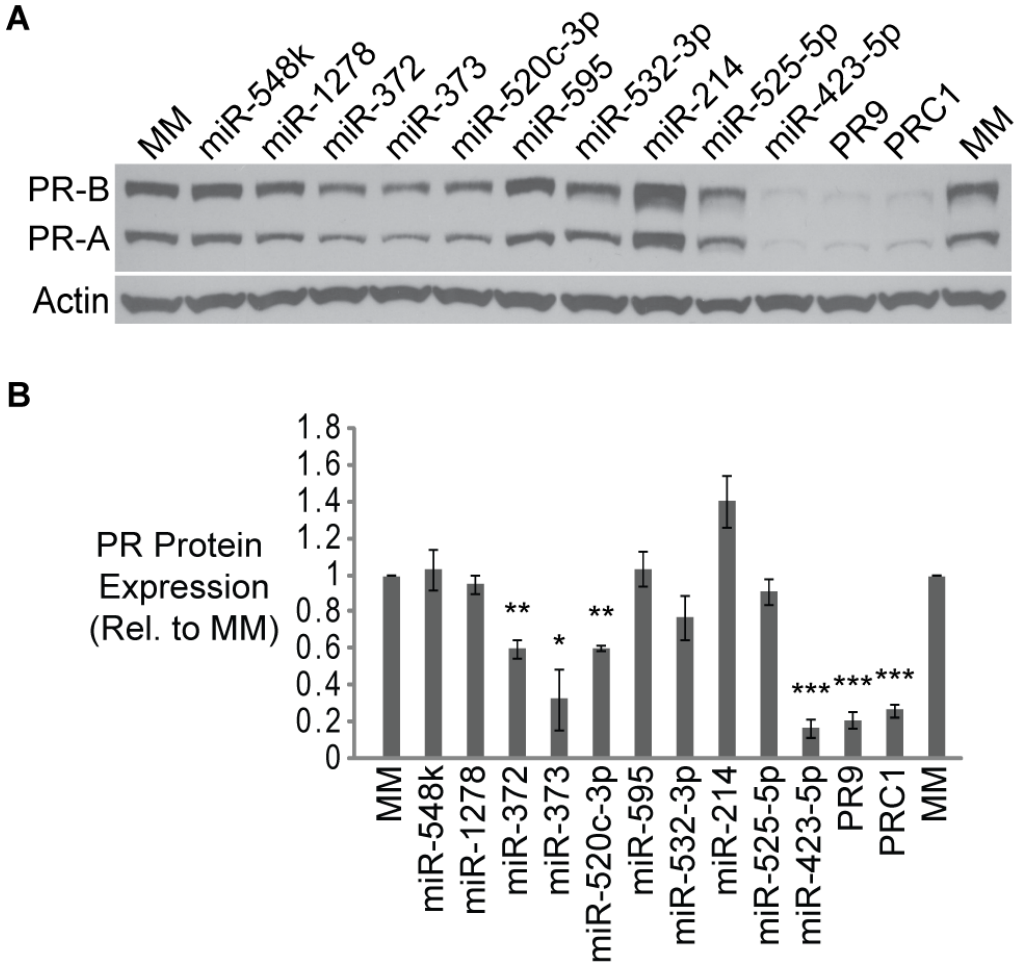


Figure 3.4. Dose dependent inhibition of PR by miR-423-5p. (A) Western analysis showing inhibition of PR protein expression by miR-423-5p. (B) Quantification of independent dose response experiments measuring reduction of PR protein levels by miR-423-5p. Error bars indicate s.d. (n = 4).

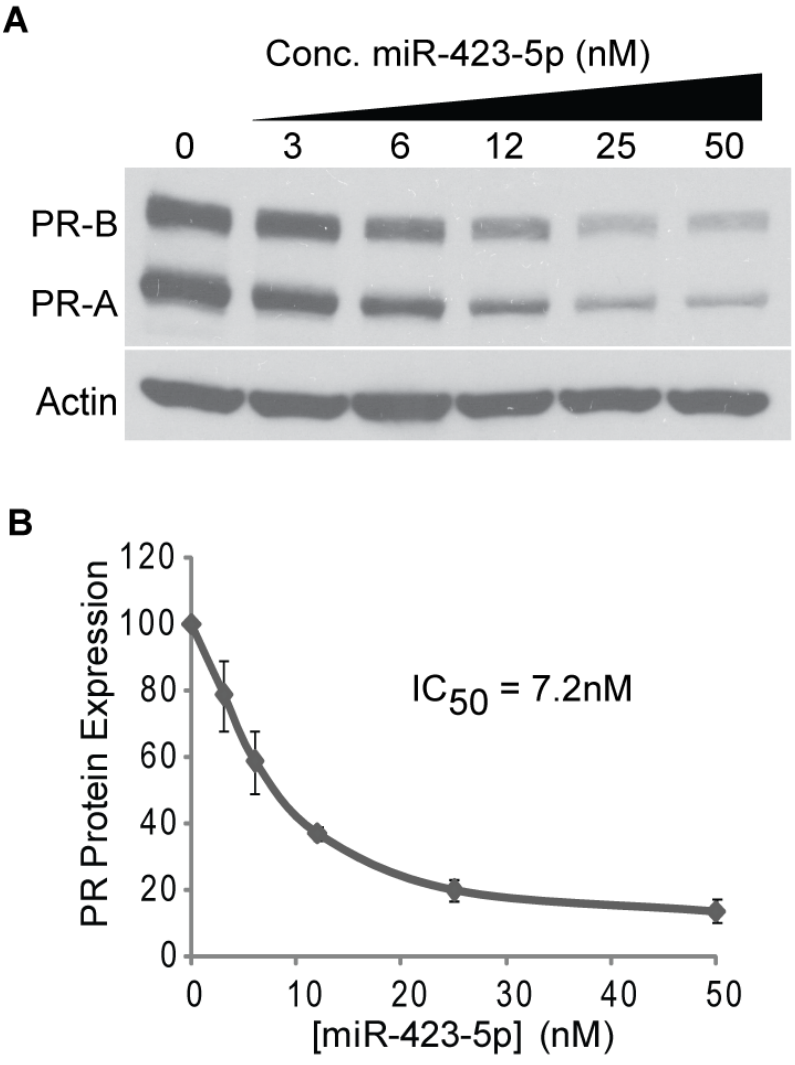


Figure 3.5. miR-423-5p inhibits PR RNA expression. (A,B) RT-qPCR showing inhibition of (A) PR mRNA and (B) PR hnRNA expression by miR-423-5p. miRNA mimics were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

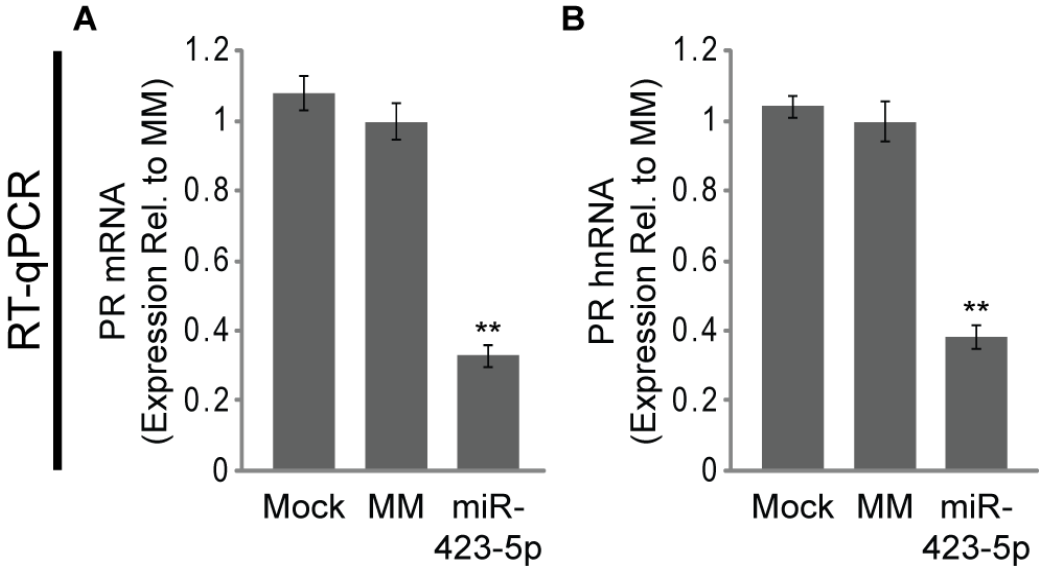


Figure 3.6. miR-423-5p inhibits PR transcription. (A,B) ChIP-qPCR analysis of (A) RNA Pol II and (B) H3K9me2 occupancy on the PR promoter following treatment with miR-423-5p. miRNA mimics were added to cells at 50 nM . Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

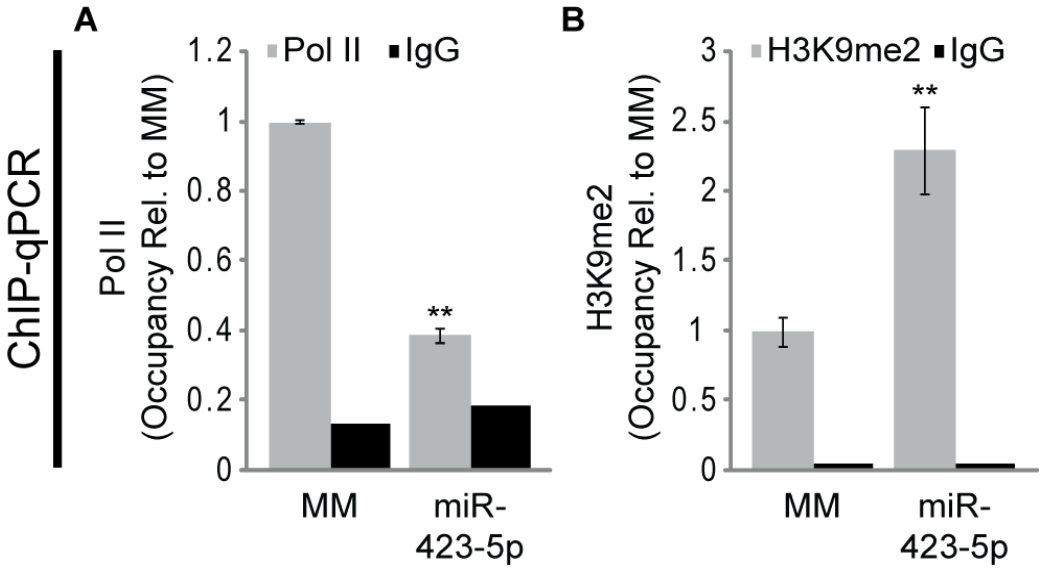


Figure 3.7. Inhibition of PR by pre-miR-423. (A) Structure of pre-miR-423 duplex (seed sequences shown in lower case). (B) RT-qPCR showing inhibition of PR mRNA by pre-miR-423. Error bars indicate s.d. (n = 3). miRNA mimics were added to cells at 25 nM. *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

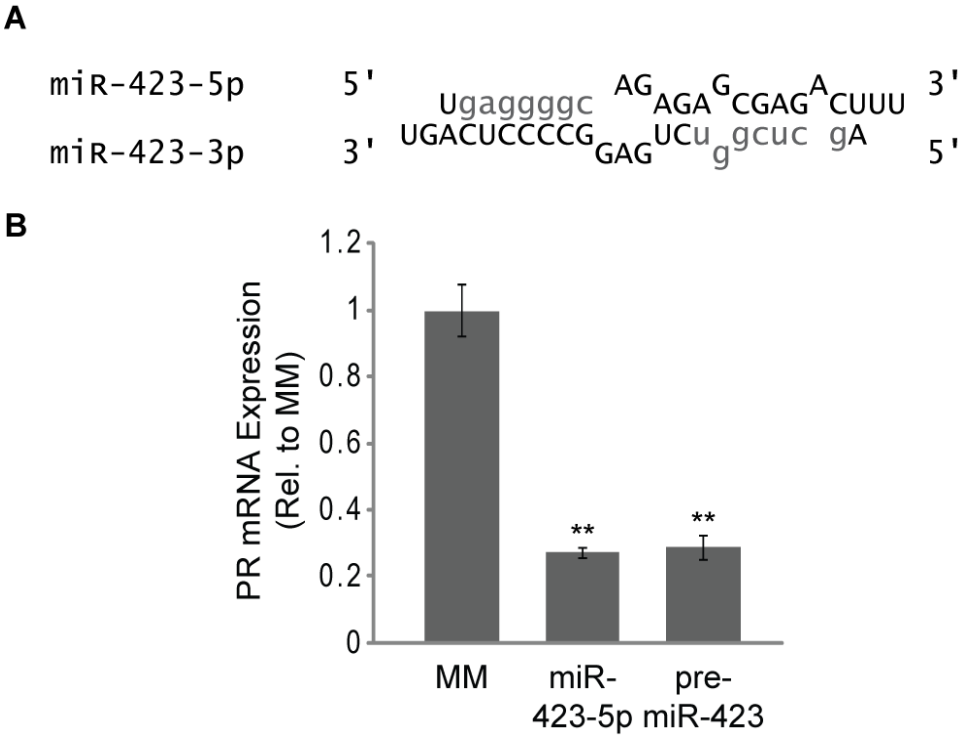


Figure 3.8. Detection of miR-423-5p and pre-miR-423 expression. (A,B) RT-qPCR analysis of whole cell (A) miR-423-5p and (B) miR-423-3p levels following treatment with miR-423-5p and pre-miR-423. (C,D) RT-qPCR analysis of nuclear (C) miR-423-5p and (D) miR-423-3p levels following treatment with miR-423-5p and pre-miR-423. miRNA mimics were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. * $P < 0.05$, ** $P < 0.01$.

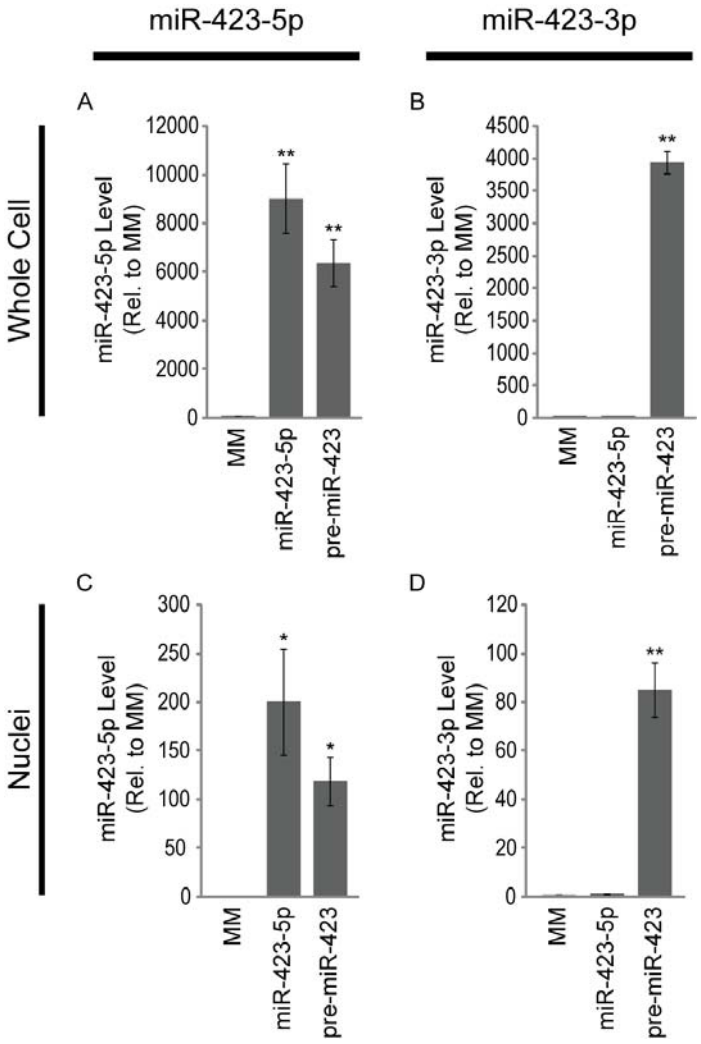


Figure 3.9. miRNAs complementary to the PR promoter inhibit PR expression. (A,B) RT-qPCR showing inhibition of (A) PR mRNA and (B) PR hnRNA expression by multiple miRNA mimics. Error bars indicate s.d. (n = 3). miRNA mimics were added to cells at 25 nM. *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

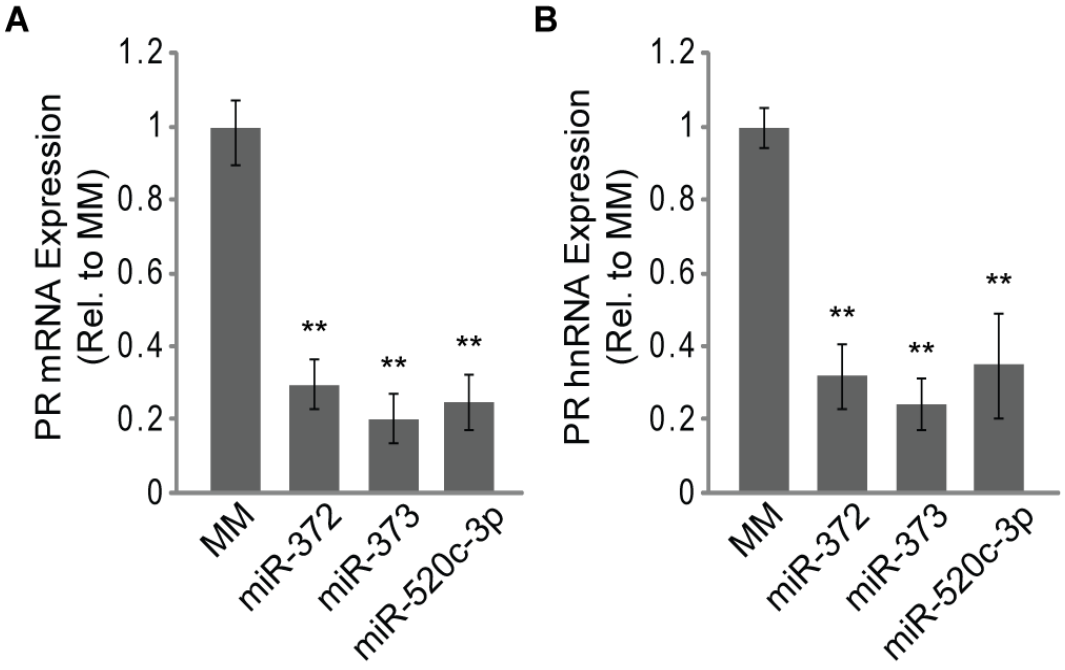


Figure 3.10. miR-423-5p is complementary to the PR promoter. (A) Alignment of miR-423-5p with ncRNA overlapping the PR promoter (seed sequence shown in lower case). (B) PhastCons conservation analysis of the miR-423-5p target site within the PR promoter (seed sequence target site underlined, sequence listed 3' to 5').

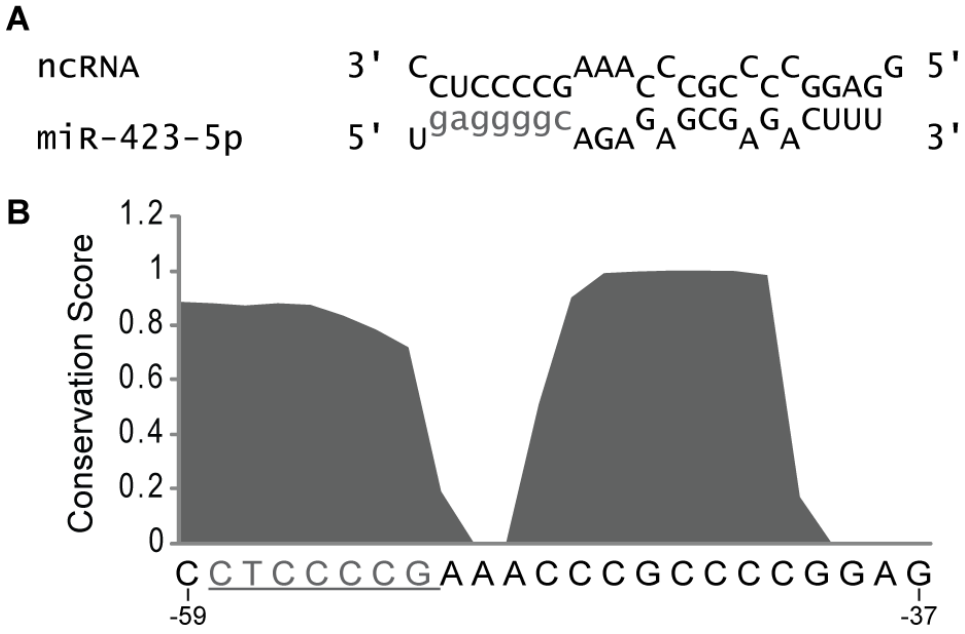


Figure 3.11. Inhibition of PR by miR-423-5p requires seed sequence complementarity. (A) Alignment of mutant miR-423-5p mimics with ncRNA overlapping the PR promoter (mismatched bases underlined). (B) RT-qPCR measuring PR mRNA expression following treatment with mutant miR-423-5p mimics. (C) Western analysis of PR expression following treatment with mutant miR-423-5p mimics. (D) Quantification of independent experiments measuring PR protein expression after treatment with mutant miR-423-5p mimics. Duplexes were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

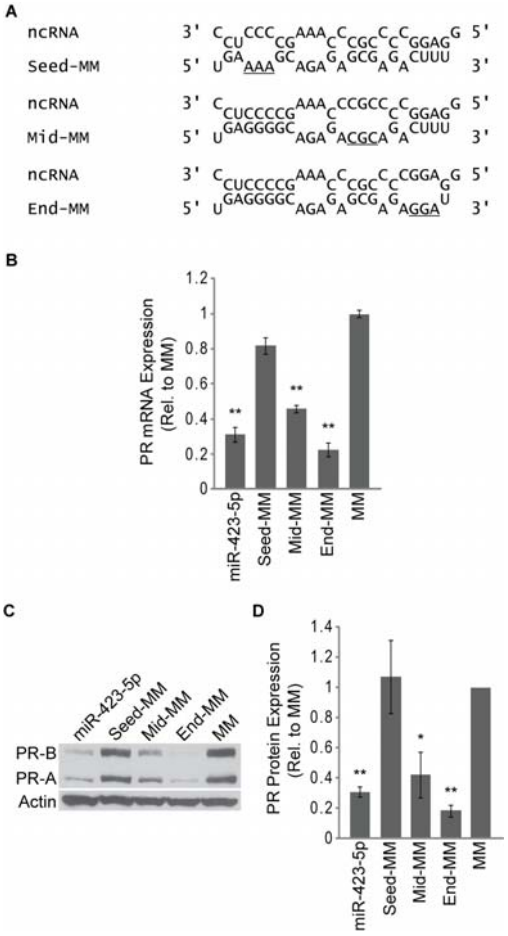


Figure 3.12. Inhibition of PR by miR-423-5p is specific to RNA. RT-qPCR measuring PR mRNA expression following treatment with a DNA analog of miR-423-5p. Duplexes were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

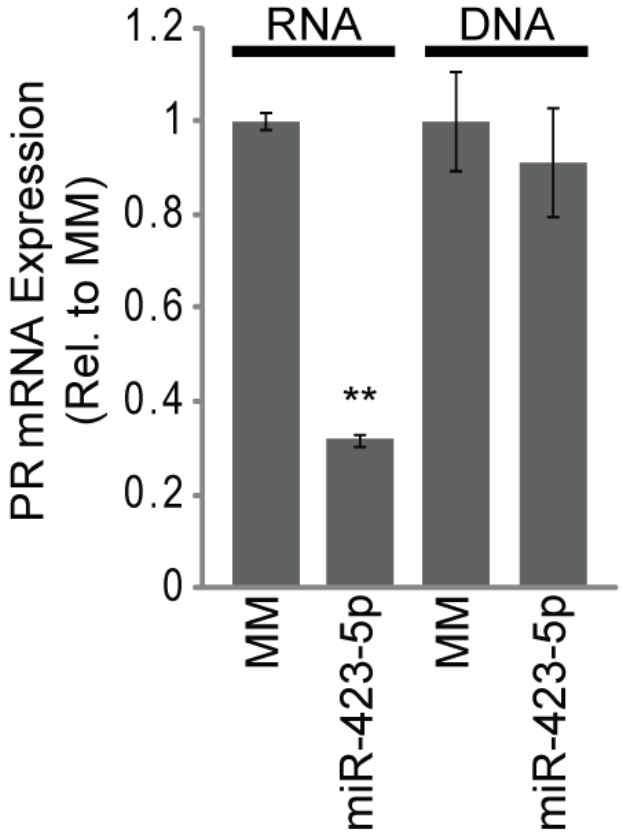


Figure 3.13. miR-423-5p-induced silencing of PR is independent of DNA methylation. (A) Methylation specific PCR of the PR promoter after treatment with miR-423-5p. U, unmethylated. M, methylated. (B) Bisulfite sequencing of the PR promoter following addition of miR-423-5p. Open circles, unmethylated CpG. Closed circles, methylated CpG. Circles represent individual CpG dinucleotides. miRNA mimics were added to cells at 25 nM.

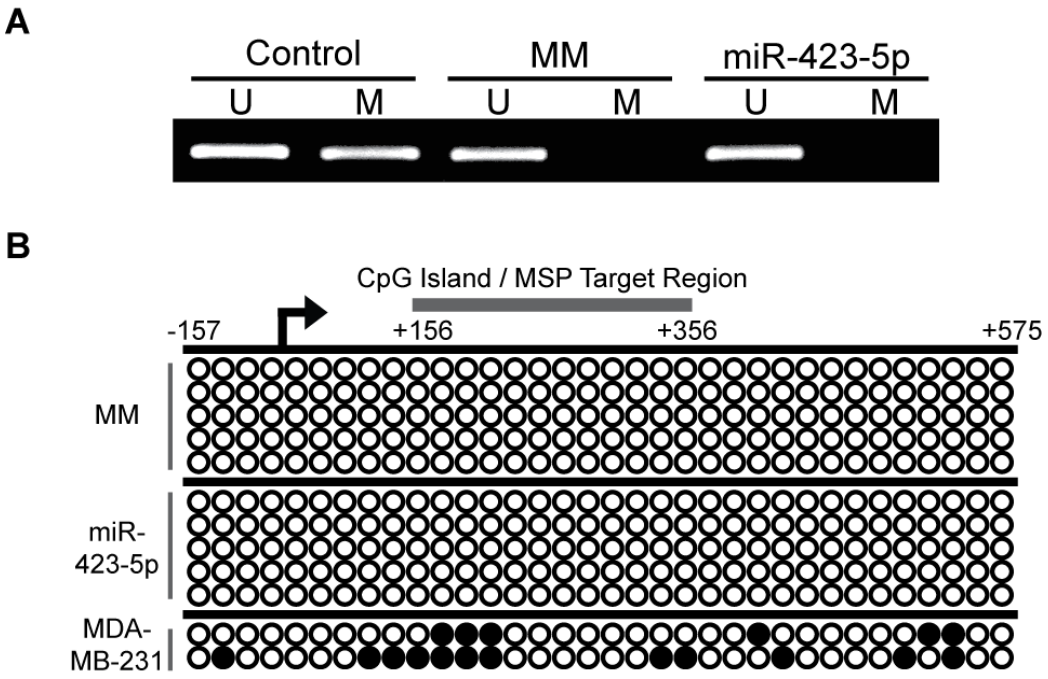


Figure 3.14. miR-423-5p-induced silencing of PR is transient. Timecourse of PR protein inhibition by miR-423-5p. miRNA mimics were added to cells at 25 nM.

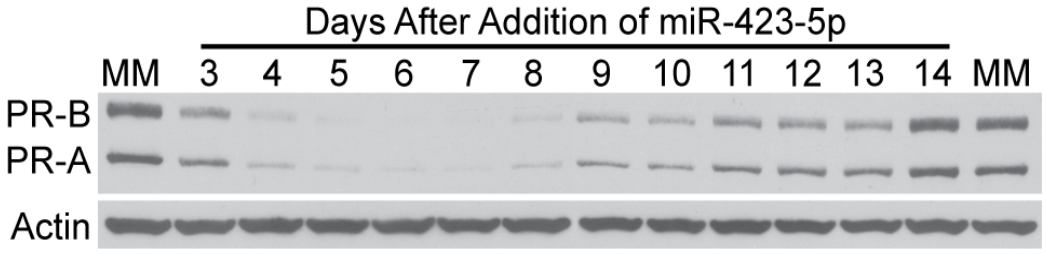


Figure 3.15. miR-423-5p recruits AGO2 to the ncRNA overlapping the PR promoter. (A,B) RIP analysis of (A) AGO1 and (B) AGO2 recruitment to the ncRNA overlapping the PR promoter by miR-423-5p. miRNA mimics were added to cells at 50 nM.

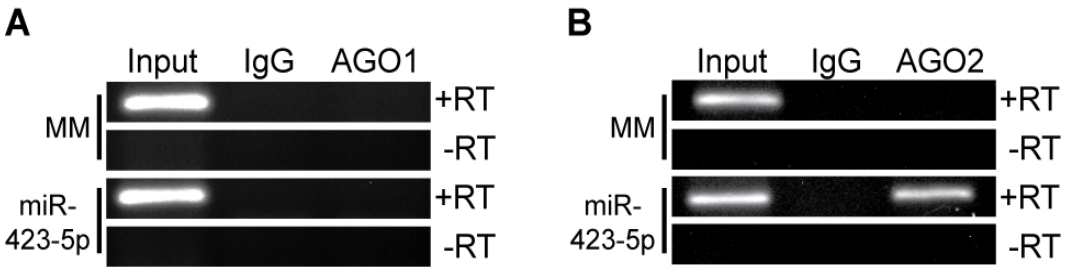


Figure 3.16. Effect of miR-423-5p addition on ncRNA expression. RT qPCR analysis of ncRNA expression following treatment with miR-423-5p. miRNA mimics were added to cells at 25 nM. Error bars indicate s.d. (n = 3).

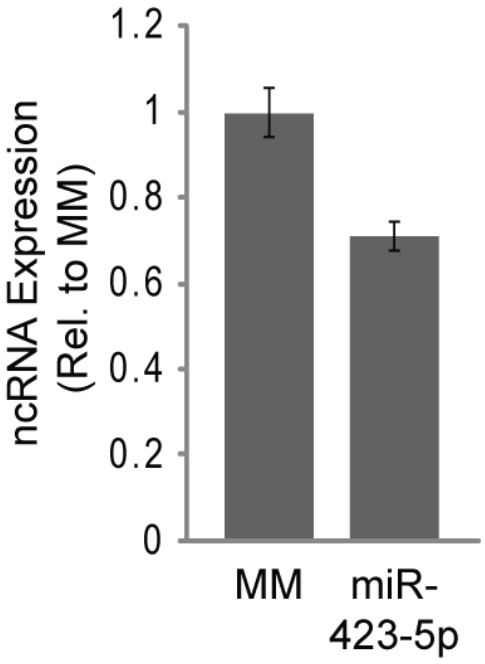


Figure 3.17. Inhibition of PR expression by miR-423-5p requires AGO2. (A) RT-qPCR analysis of AGO1 mRNA expression following treatment with siRNA against AGO1. (B) RT-qPCR analysis of PR mRNA inhibition by miR-423-5p following AGO1 knockdown. (C) RT-qPCR analysis of AGO2 mRNA expression following treatment with siRNA against AGO2. (D) RT-qPCR analysis of PR mRNA inhibition by miR-423-5p following AGO2 knockdown. TF1 = transfection 1, TF2 = transfection 2. siRNAs and miRNA mimics were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

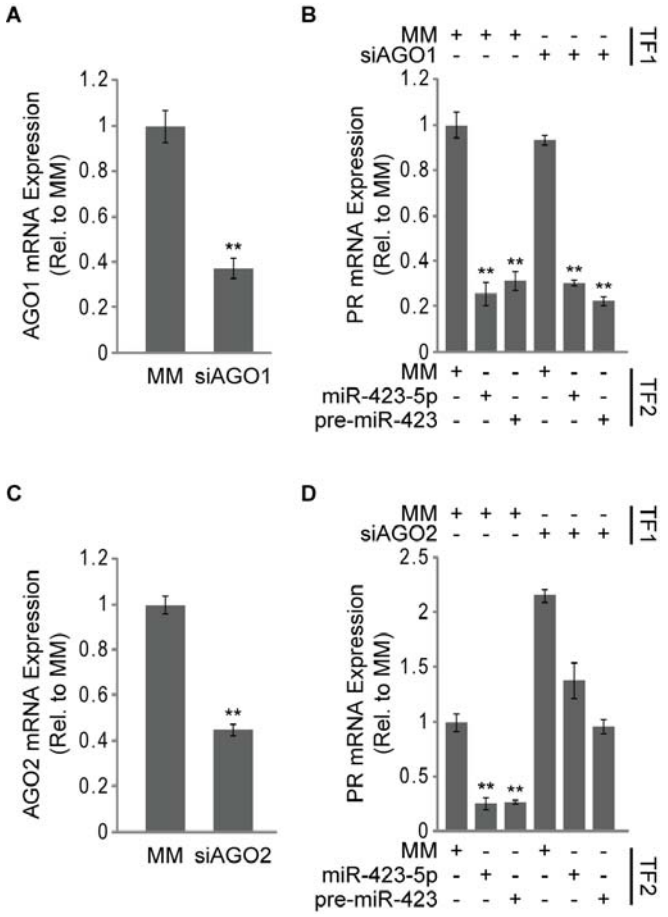


Figure 3.18. Effect of miR-423-5p inhibition on PR expression. (A) miR-423-5p expression in T47D cells following treatment with miRNA inhibitors. (B) PR mRNA expression following treatment with miRNA inhibitors in T47D cells. (C) miR-423-5p expression in MCF7 cells following treatment with miRNA inhibitors. (D) PR mRNA expression following treatment with miRNA inhibitors in MCF7 cells. Error bars indicate s.d. (n = 3). 2'-O-methyl RNAs were added to cells at 100 nM. *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. * $P < 0.05$, *** $P < 0.001$.

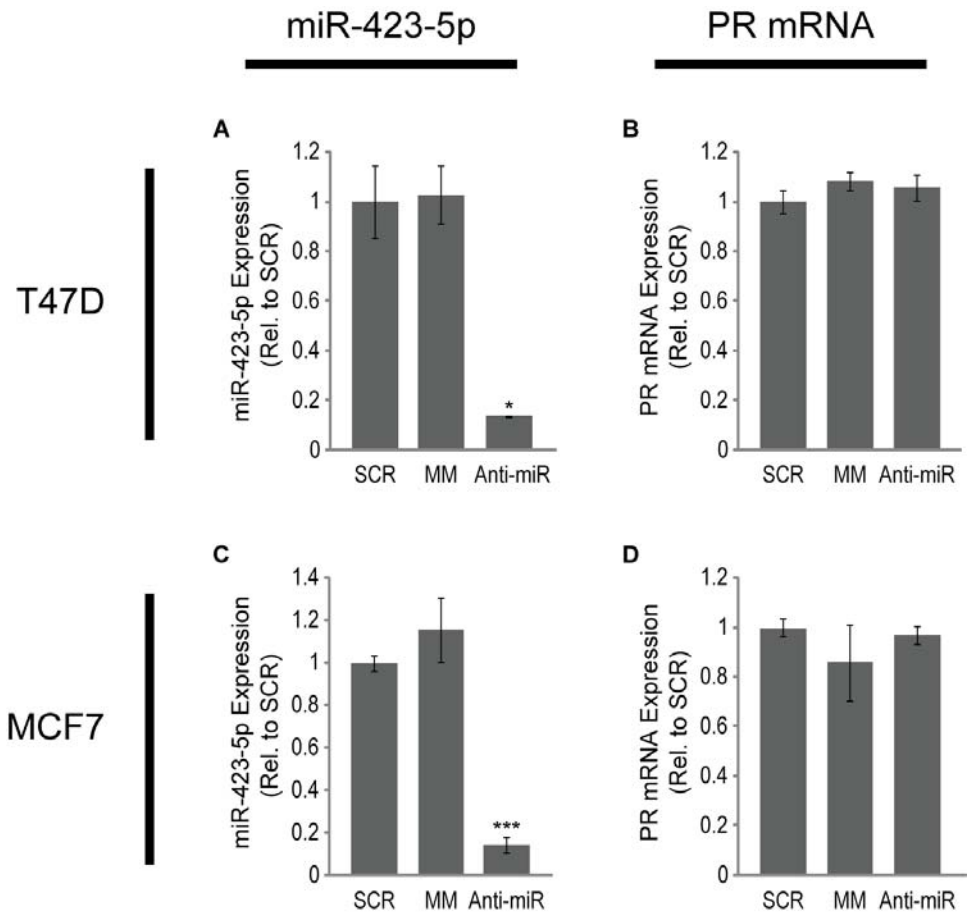


Figure 3.19. Inhibition of PR expression by miR-423-5p in MCF7 cells. (A) Western analysis of PR expression following treatment with miR-423-5p in MCF7 cells. (B,C) RT-qPCR showing inhibition of (B) PR mRNA and (C) PR hnRNA expression by miR-423-5p in MCF7 cells. Error bars indicate s.d. (n = 3). miRNA mimics were added to cells at 25 nM. *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

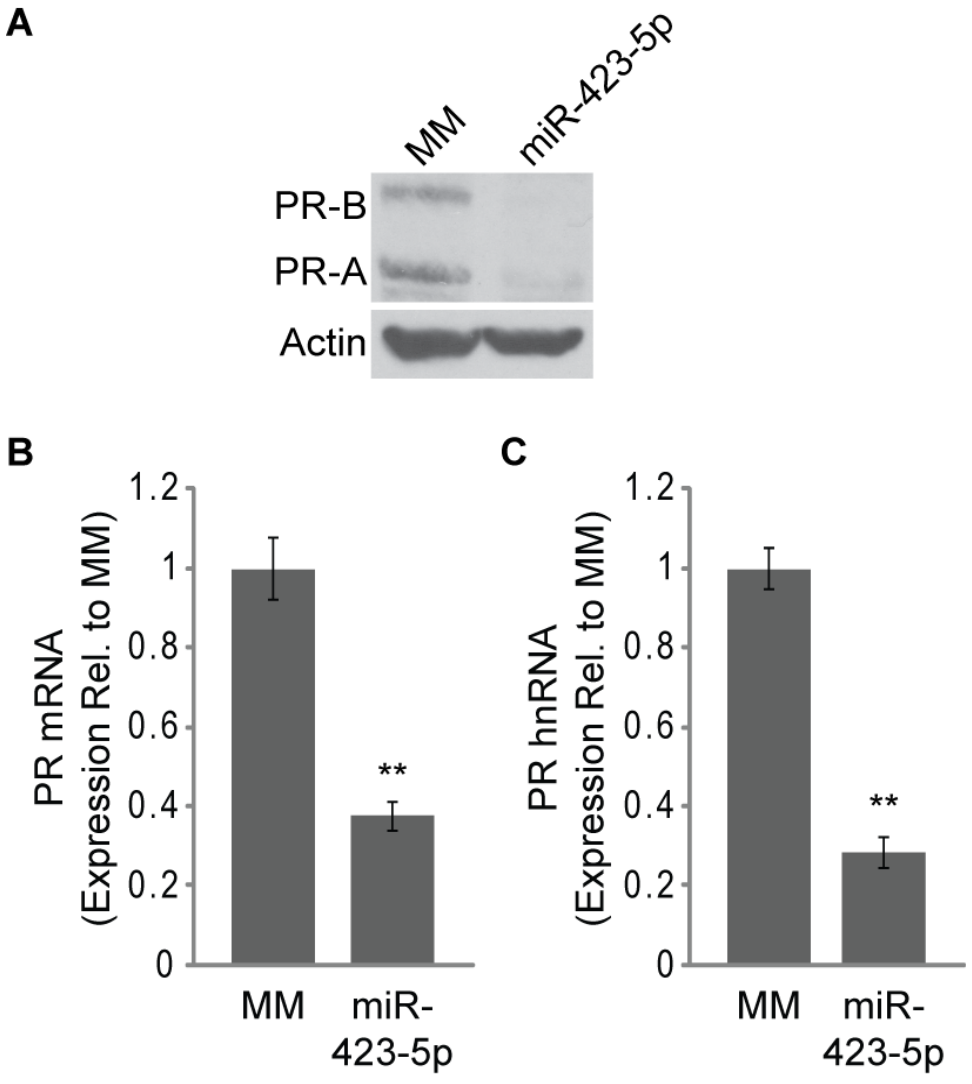


Figure 3.20. Inhibition of IGSF1 expression by miR-423-5p. (A) PhastCons conservation analysis of the miR-423-5p target site within the IGSF1 promoter (seed sequence target site underlined, sequence listed 3' to 5'). (B) RT-PCR detection of IGSF1 promoter-overlapping ncRNA and mRNA expression. (C,D) RT-qPCR showing inhibition of (C) IGSF1 mRNA and (D) IGSF1 hnRNA expression by miR-423-5p. miRNA mimics were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

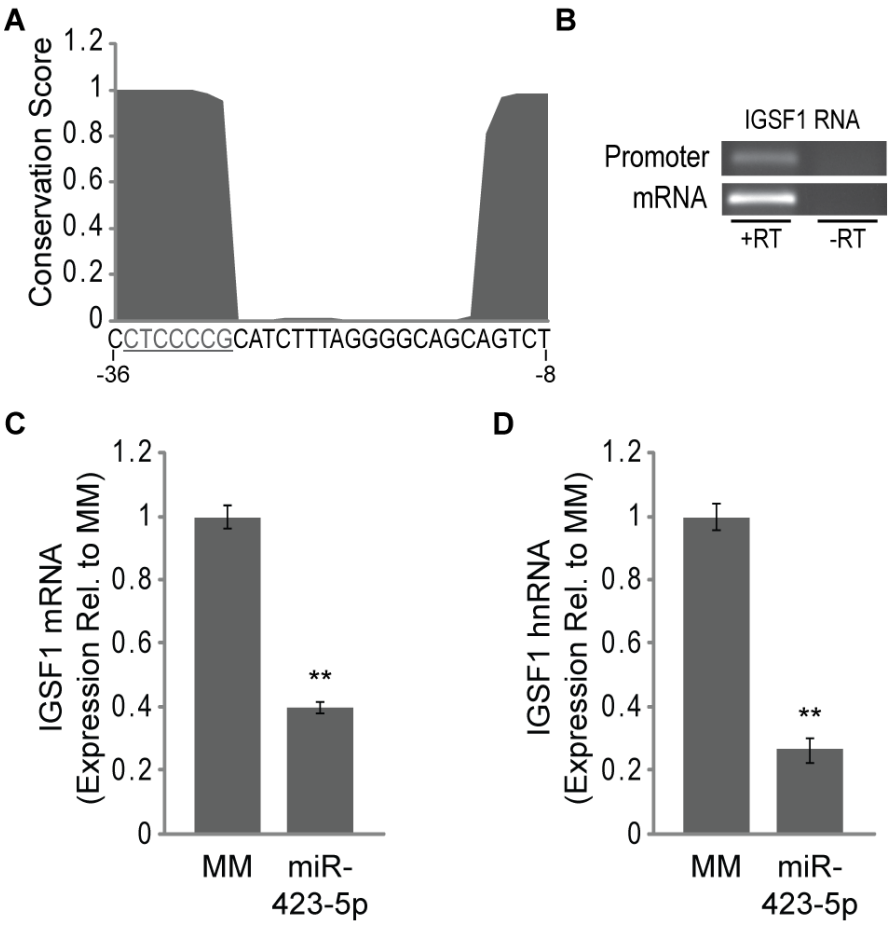


Figure 3.21. Inhibition of IGSF1 expression by miR-423-5p requires seed sequence complementarity. (A) Alignment of miR-423-5p and mutant miR-423-5p mimics with ncRNA overlapping the IGSF1 promoter (seed sequence show in lower case, mismatched bases underlined). (B) RT-qPCR analysis of IGSF1 mRNA expression following treatment with mutant miR-423-5p mimics. Duplexes were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

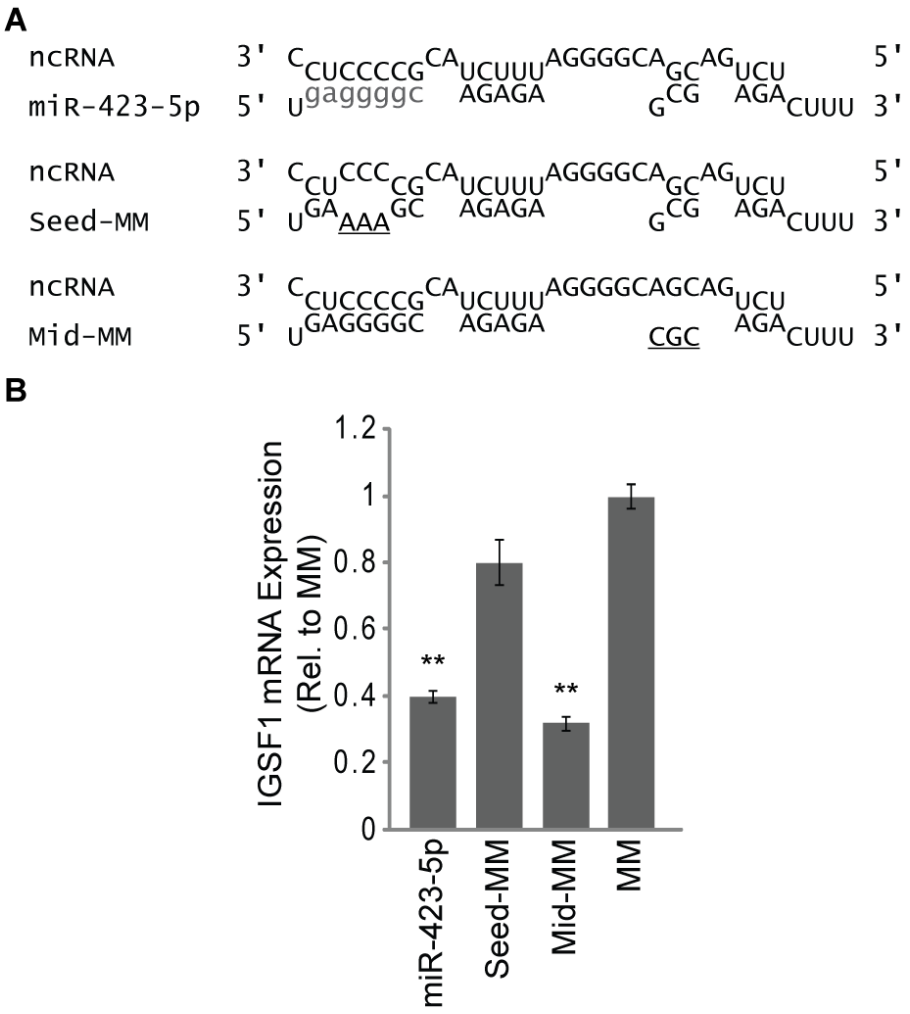
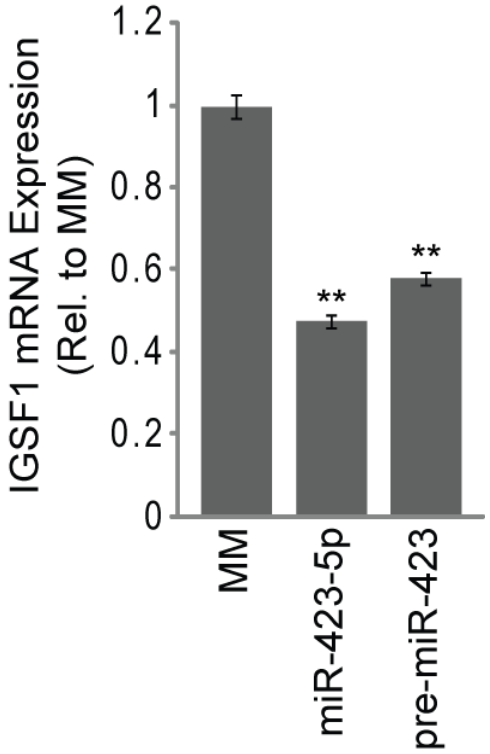


Figure 3.22. Inhibition of IGSF1 expression by pre-miR-423. RT-qPCR analysis of IGSF1 mRNA expression following treatment with pre-miR-423 mimic. miRNA mimics were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.



References For Chapter 3

1. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289-1292 (2004).
2. Suzuki, K. *et al.* Prolonged transcriptional silencing and CpG methylation induced by siRNAs targeted to the HIV-1 promoter region. *J RNAi Gene Silencing* **1**, 66-78 (2005).
3. Ting, A. H., Schuebel, K. E., Herman, J. G. & Baylin, S. B. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat Genet* **37**, 906-910 (2005).
4. Janowski, B. A. *et al.* Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat Chem Biol* **1**, 216-222 (2005).
5. Pulukuri, S. M. & Rao, J. S. Small interfering RNA directed reversal of urokinase plasminogen activator demethylation inhibits prostate tumor growth and metastasis. *Cancer Res* **67**, 6637-6646 (2007).
6. Napoli, S., Pastori, C., Magistri, M., Carbone, G. M. & Catapano, C. V. Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J* **28**, 1708-1719 (2009).
7. Hawkins, P. G., Santoso, S., Adams, C., Anest, V. & Morris, K. V. Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. *Nucleic Acids Res* **37**, 2984-2995 (2009).
8. Li, L. C. *et al.* Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci U S A* **103**, 17337-17342 (2006).
9. Janowski, B. A. *et al.* Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol* **3**, 166-173 (2007).
10. Chen, Z. *et al.* Antitumor effect of dsRNA-induced p21(WAF1/CIP1) gene activation in human bladder cancer cells. *Mol Cancer Ther* **7**, 698-703 (2008).
11. Huang, V. *et al.* RNAa is conserved in mammalian cells. *PLoS One* **5**, e8848 (2010).
12. Morris, K. V., Santoso, S., Turner, A. M., Pastori, C. & Hawkins, P. G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet* **4**, e1000258 (2008).
13. Han, J., Kim, D. & Morris, K. V. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci U S A* **104**, 12422-12427 (2007).
14. Schwartz, J. C. *et al.* Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol* **15**, 842-848 (2008).
15. Morris, K. V. RNA-directed transcriptional gene silencing and activation in human cells. *Oligonucleotides* **19**, 299-306 (2009).
16. Janowski, B. A. *et al.* Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* **13**, 787-792 (2006).
17. Kim, D. H., Villeneuve, L. M., Morris, K. V. & Rossi, J. J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **13**, 793-797 (2006).
18. Chu, Y., Yue, X., Younger, S. T., Janowski, B. A. & Corey, D. R. Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-

coding transcript at the progesterone receptor promoter. *Nucleic Acids Res* **38**, 7736-7748 (2010).
19. Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16**, 2733-2742 (2002).
20. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716-719 (2003).
21. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833-1837 (2002).
22. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672-676 (2004).
23. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233 (2009).
24. Ambros, V. The evolution of our thinking about microRNAs. *Nat Med* **14**, 1036-1040 (2008).
25. Place, R. F., Li, L. C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* **105**, 1608-1613 (2008).
26. Kim, D. H., Saetrom, P., Snove, O., Jr. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 16230-16235 (2008).
27. Younger, S. T., Pertsemlidis, A. & Corey, D. R. Predicting potential miRNA target sites within gene promoters. *Bioorg Med Chem Lett* **19**, 3791-3794 (2009).
28. Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232-2237 (2002).
29. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-144 (2006).
30. Ebralidze, A., Wang, Y., Petkova, V., Ebralidse, K. & Junghans, R. P. RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science* **303**, 383-387 (2004).
31. John, B. *et al.* Human MicroRNA targets. *PLoS Biol* **2**, e363 (2004).
32. Gingeras, T. R. Origin of phenotypes: genes and transcripts. *Genome Res* **17**, 682-690 (2007).
33. Yue, X. *et al.* Transcriptional regulation by small RNAs at sequences downstream from 3' gene termini. *Nat Chem Biol* **6**, 621-629 (2010).

Chapter 4: Transcriptional Regulation by miRNAs that target sequences downstream of gene termini

4.1 Introduction

Recent studies have revealed unexpectedly complex transcription within the human genome. These investigations have shown that many protein-coding genes are flanked at both their 5' and 3' termini by noncoding (ncRNA) transcripts[1-2]. These transcripts are produced in both the sense and antisense direction with respect to the adjacent gene. The natural biological function of these ncRNAs has not been determined, however their proximity to protein-coding regions suggest they might be involved in regulating gene expression.

Consistent with reports of complex transcription at gene loci, we have identified ncRNA transcripts that overlap either the promoter or terminus of the human progesterone receptor (PR) gene[3-4]. We have shown that small duplex RNAs that are fully complementary to these transcripts can modulate transcription of the PR gene in a potent and robust fashion[3-8]. Our experiments have demonstrated that these small RNAs recruit argonaute 2 (AGO2) to their respective target ncRNAs and alter levels of RNA Pol II at the PR gene transcription start site[3-5,7].

Initially, modulation of transcription by small RNAs complementary to sequences beyond the 3' terminus of the PR gene was puzzling because approximately 100,000 bases separate the target site from the gene's transcription start site. Experiments have revealed, however, that the PR locus loops and juxtaposes the transcription start site with the gene terminus[4]. This looping occurs in multiple cell lines

independent of whether the PR gene is expressed at high or low levels, a modulatory small RNA is present, or biological stimuli are applied to cells. The proximity of the 5' and 3' gene termini provides a direct path allowing transcriptional modulation by small RNAs targeted to distal genomic regions.

MicroRNAs (miRNAs) are endogenous small RNAs that can recognize sequences within 3' UTRs of mRNA transcripts[9-10]. In addition, recent reports have shown that miRNAs can target ncRNAs that overlap gene promoters and subsequently regulate transcription[11-14]. Our observation of efficient gene regulation by duplex RNAs that are fully complementary to sequences downstream of gene termini suggests that miRNAs might also target these sequences, although recognition could occur through partial complementarity to their target sequences. To test this hypothesis we investigated the genome-wide occurrence of miRNA target sites within sequences immediately downstream of protein-coding loci and tested whether miRNAs complementary to the well-characterized ncRNA overlapping the 3' terminus of the PR gene could modulate gene expression.

4.2 Significance

Prior to the work presented in this chapter, one report had been published describing synthetic small duplex RNAs that were capable of targeting regions beyond gene termini and regulating transcription of the upstream gene[4]. In this chapter we reproduce those results using miRNA mimics, demonstrating that endogenous small RNA sequences are also capable of targeting gene termini. Together with the findings from the preceding chapters, our observations here strengthen the argument that miRNAs are capable of

directly regulating transcription. Furthermore, they highlight a more diverse role for miRNAs in regulating gene expression than has previously been appreciated.

4.3 Detecting miRNA target sites within sequences downstream of gene termini

To evaluate the potential for miRNAs to target the 3'-termini of protein-coding genes we developed a miRNA target prediction algorithm. A major determinant of miRNA targeting is perfect base complementarity between the miRNA seed sequence (bases 2-8 of the mature miRNA) and its target sequence. Our algorithm searches for miRNA seed sequence matches and, if found, calculates additional metrics for comparison (**Figure 4.1A**). First, we determine the minimum free energy of hybridization (MFE) between the miRNA and its target sequences. Second, we calculate a complementarity score using the Needleman-Wunsch algorithm with a scoring matrix optimized for small RNA interactions[15]. While these values are not mutually exclusive, they provide two independent criteria for comparison.

Our algorithm incorporates several additional layers of sequence analysis for predicted miRNA targets. First, we discriminate unique target sequences from those that occur within repeat elements. Next, we determine cross-species conservation values for target sequences using the PhastCons analysis of the multiz alignment of 44 vertebrate species[16]. Finally, we analyze the predicted RNA secondary structure of target sequences calculated using EvoFold[17].

miRNA sequences were obtained from miRBase, the public repository for miRNAs[18]. Gene termini sequences were obtained from the UCSC genome browser and consisted of the 1000 bases immediately downstream of the annotated transcription

termination site for each gene in the human genome. We used the genomic sequences of gene termini to construct sequence sets corresponding to potential ncRNA transcripts in both the sense and antisense orientations relative to their upstream gene (**Figure 4.1B**). We also obtained sequences corresponding to the 3'-UTR of all annotated genes to compare the frequencies of matches within these datasets.

4.4 Sequences downstream of gene termini are enriched with putative miRNA target sites.

We used our algorithm to perform a genome-wide evaluation of potential miRNA target sites within putative ncRNAs produced from sequences downstream of gene termini. In addition, we compared these results to predicted target sites within 3' UTRs. We calculated the total number of seed sequence matches within each dataset and normalized the values to the number of bases within each respective dataset. We found that the frequencies of predicted miRNA target sites within downstream ncRNAs, in both the sense and antisense orientations, were indistinguishable from those within 3'-UTRs (**Figure 4.2A**).

To determine if ncRNAs overlapping gene termini are enriched with predicted miRNA target sites we compared the frequency of seed matches in our initial analysis to the frequency with which they occur in randomized sequences. Each sequence in the dataset was subjected to 100 iterations of randomization followed by screening for the number of seed sequence matches. We observed a significant enrichment of potential miRNA target sites within both sense and antisense oriented ncRNAs with respect to

randomized sequences (**Figure 4.2B**). This enrichment was comparable to that of predicted miRNA target sites within 3' UTRs (**Figure 4.2B**).

We identified many miRNAs with a striking degree of complementarity to their predicted target sites. A selected subset of predicted miRNA target sites within both sense and antisense oriented ncRNAs are shown in **Figure 4.3**. Each alignment is accompanied by the location of the target site with respect to the transcription termination site for the adjacent gene along with the MFE value for the alignment. Taken together, these results suggest that ncRNAs produced from sequences downstream of gene termini are promising candidates for miRNA targets.

4.5 miRNA target sites within the ncRNA overlapping the PR gene terminus

Our computational results prompted us to experimentally evaluate the possibility that miRNAs can target ncRNAs produced from sequences downstream of gene termini. We have previously used the human PR gene as a model system for studying transcriptional regulation by small RNAs. While characterizing the transcriptional landscape of the PR locus in our prior studies we identified a 3.2 kb ncRNA transcript that overlaps the 3'-terminus of the gene and is expressed at ~4% relative to PR mRNA[4]. The ncRNA is transcribed in the same direction as the PR gene and shares 1.7 kb of sequence content with PR mRNA. The remaining 1.5 kb of the transcript is unique to the ncRNA. Several assays including RT-PCR, rapid amplification of cDNA ends (RACE), and branched DNA (bDNA) verified that the ncRNA is not an extension of PR mRNA.

Our algorithm identified 84 putative miRNA target sites within the unique region of the ncRNA. To test the hypothesis that miRNAs can regulate gene expression

by recognizing sequences downstream of gene termini, we selected a subset of 5 miRNAs that displayed a high degree of complementarity to the target ncRNA (**Figure 4.4A, B**). We designed miRNA mimics corresponding to each miRNA that consisted of the miRNA sequence and a fully complementary RNA carrier strand. The miRNA mimics were transfected into T47D breast cancer cells and PR protein expression was monitored. An siRNA targeting PR mRNA (PRC1) and a duplex RNA previously shown to target the terminus-overlapping transcript and inhibit PR expression (PR13580) were used as positive controls for modulation of PR expression. A mismatched duplex RNA (MM) that does not affect PR expression was used as a negative control to demonstrate the specificity for recognition of the target ncRNA.

Of the 5 miRNA mimics tested, we observed that the miR-193b mimic was the most promising inhibitor of PR protein expression (**Figure 4.5**). miR-193b has a single predicted target site within the target ncRNA roughly 1000 bases downstream of the transcription termination site of the PR gene and no target sites within PR mRNA. Inhibition of PR by miR-193b was potent and dose dependent, with an IC_{50} near 15 nM (**Figure 4.6**), a value similar to the previously-determined IC_{50} for PRC1 (11.5 nM) or PR13580 (10.7 nM).

4.6 Inhibition of transcription by miRNAs that target the ncRNA overlapping the PR terminus

To determine if inhibition of PR was occurring at the level of RNA we used quantitative RT-PCR (RT-qPCR) to evaluate expression of PR mRNA. We observed a 70% reduction in PR mRNA expression after treatment with miR-193b which was similar to the level

achieved by fully complementary RNA PR13580 (**Figure 4.7A**). Designed inhibitory small RNAs that are fully complementary to the PR terminus-overlapping ncRNA also decrease expression of the target ncRNA. Using RT-qPCR we found that miR-193b decreased expression of the ncRNA by 70%, which again was similar to the levels observed following treatment with PR13580 (**Figure 4.7B**).

To investigate whether inhibition of PR was occurring at the level of transcription we used two independent assays. First, we used RT-qPCR to monitor expression of pre-spliced PR mRNA, also termed heteronuclear RNA (hnRNA). We observed a 50% decrease in PR hnRNA expression following treatment with miR-193b (**Figure 4.8A**). Second, we used chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR) to measure RNA Polymerase II (Pol II) occupancy on the PR promoter. Treatment with miR-193b decreased Pol II occupancy by greater than 80% (**Figure 4.8B**). Taken together, these results indicate that miR-193b is capable of targeting the ncRNA overlapping the PR gene terminus and regulating transcription of the PR gene.

4.7 Conclusion

miRNAs have typically been associated with recognition of mRNA transcripts, usually at sequences within 3'-UTRs. miRNA binding results in destabilization of the target mRNA and ultimately reduction in target gene expression[19]. It has become undoubtedly clear that miRNAs are powerful regulators of gene expression at the post-transcriptional level.

Transcriptome profiling studies, however, have transformed our current understanding of cellular RNAs. These studies have found that many regions of the genome that do not encode protein still produce RNA transcripts which are termed noncoding RNAs (ncRNAs). While some of these ncRNAs are directly involved in transcriptional regulation, the molecular function of most ncRNAs remains unclear[20-27]. In addition, it has been proposed that certain classes of ncRNAs may be promising targets for the development of therapeutics[28]. Regardless of ncRNA function, it does not appear that miRNAs can discriminate their targets based on protein coding potential and it is reasonable to hypothesize that many ncRNAs may be targeted by miRNAs[29].

These transcriptome profiling studies have revealed that promoter regions of protein-coding loci are overlapped by ncRNA transcripts[1-2]. These ncRNAs could serve as miRNA targets and the fact that they overlap key regulatory regions of the genome suggests that recognition by miRNAs may affect gene expression. Indeed, several reports have characterized miRNAs that target promoter-overlapping ncRNAs and subsequently regulate transcription of the downstream gene[11-13]. The ability of miRNAs to target gene promoters not only increases the sequence space available for targeting, it also complements the well-known role of miRNAs in post-transcriptional gene regulation.

Robust transcriptional regulation by miRNAs that target ncRNAs overlapping gene promoters prompted us to search for additional classes of ncRNAs that might serve as targets for small RNAs. Specifically, we focused on ncRNAs that overlap gene termini. Terminus-overlapping ncRNAs, similar to promoter-overlapping RNAs, are a common feature of protein-coding loci and their proximity to protein-coding genes also suggests that they may play a role in gene regulation. Furthermore, using the PR gene as

a model system we have been able to identify small duplex RNAs that are fully complementary to a ncRNA that overlaps the PR gene terminus and are capable of regulating PR transcription.

In this study we have found that putative ncRNA transcripts that overlap gene termini are significantly enriched with predicted miRNA target sites. In addition, we have experimentally tested a subset of miRNAs that display a high degree of complementarity to a ncRNA overlapping the PR gene terminus and found that miR-193b is a potent inhibitor of PR expression. In contrast to our previous studies with designed small RNAs where introduction of mismatched bases disrupted silencing activity, miR-193b is extensively mismatched relative to its target sequence. In addition, our designed small RNAs targeted just over 500 bases downstream of the PR gene while miR-193b targets almost 1000 bases downstream.

Our findings extend miRNA function to the recognition of sequences downstream of gene termini and also highlight a largely unappreciated role for gene termini in transcriptional regulation. However, miRNAs that target gene termini must overcome some mechanistic hurdles before they can regulate transcription of the upstream gene. For example, at the PR locus more than 100 kb separates the promoter from the gene's terminus. Experiments have revealed, however, that the PR promoter and terminus are in close physical proximity. Such gene looping is relatively common and may provide a scaffold for communication between miRNAs that target gene termini and the transcription machinery at the promoter of the upstream gene (**Figure 4.9**)[30-32].

As more comprehensive 3-dimensional maps of the human genome are compiled, it will be interesting to evaluate the ability of miRNAs to regulate transcription of a given

gene when targeted to more distal or even inter-chromosomal regions of the genome[33]. The ability to integrate signals over large genomic distances raises the possibility that miRNAs may play a role in long-range chromatin interactions and overall genome structure.

Figure 4.1. Computational approach for identifying miRNAs that target gene termini. (A) Schematic of algorithm used to identify potential miRNA target sites. (B) Diagram of sequences analyzed for miRNA target sites.

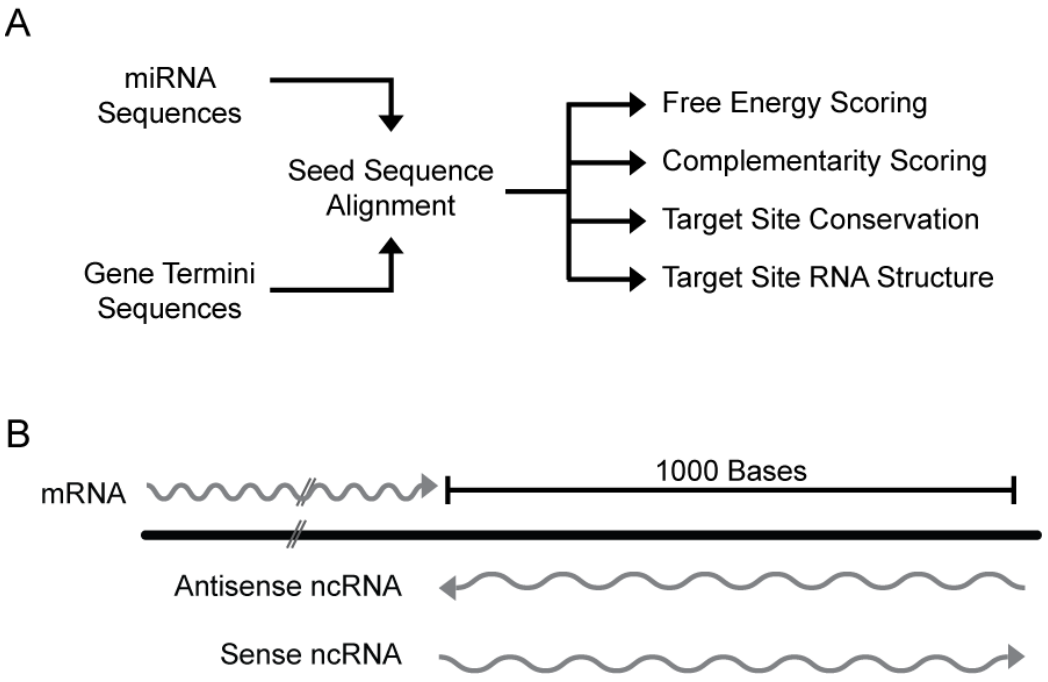


Figure 4.2. Regions beyond gene termini are enriched with putative miRNA target sites. (A) Relative frequencies of seed sequence matches within 3'UTRs and sequences downstream of gene termini. (B) Enrichment of seed sequences matches within 3'UTRs and sequences downstream of gene termini with respect to randomized sequences. ** $P < 0.01$.

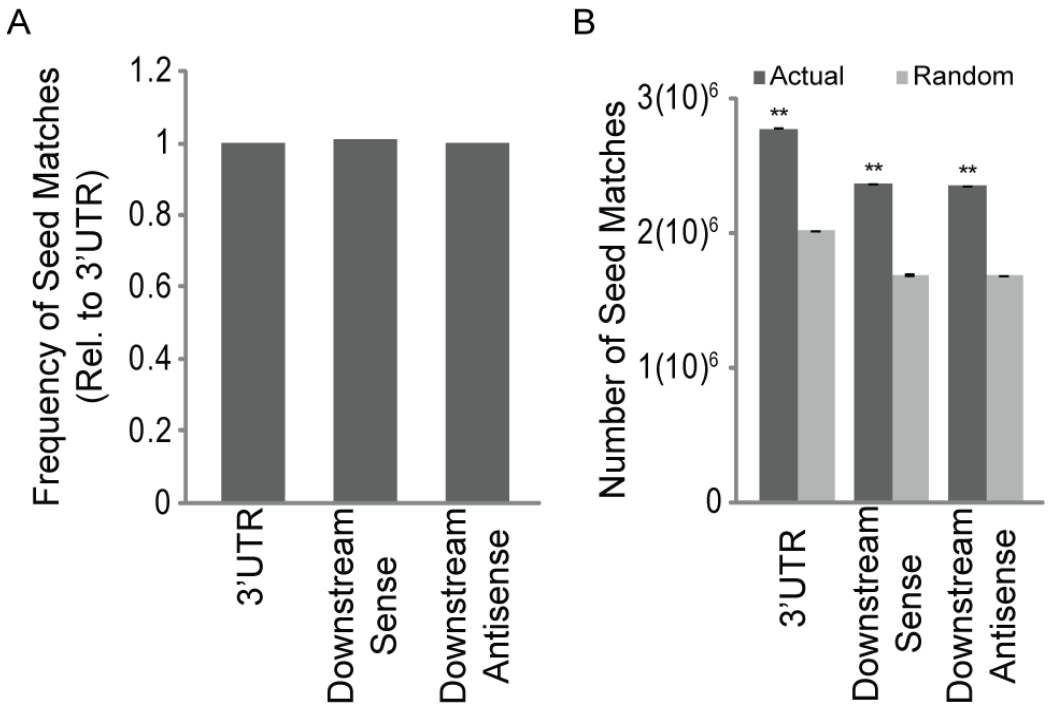


Figure 4.3. Examples of miRNAs that are highly complementary to regions downstream of gene termini (seed sequences show in red).

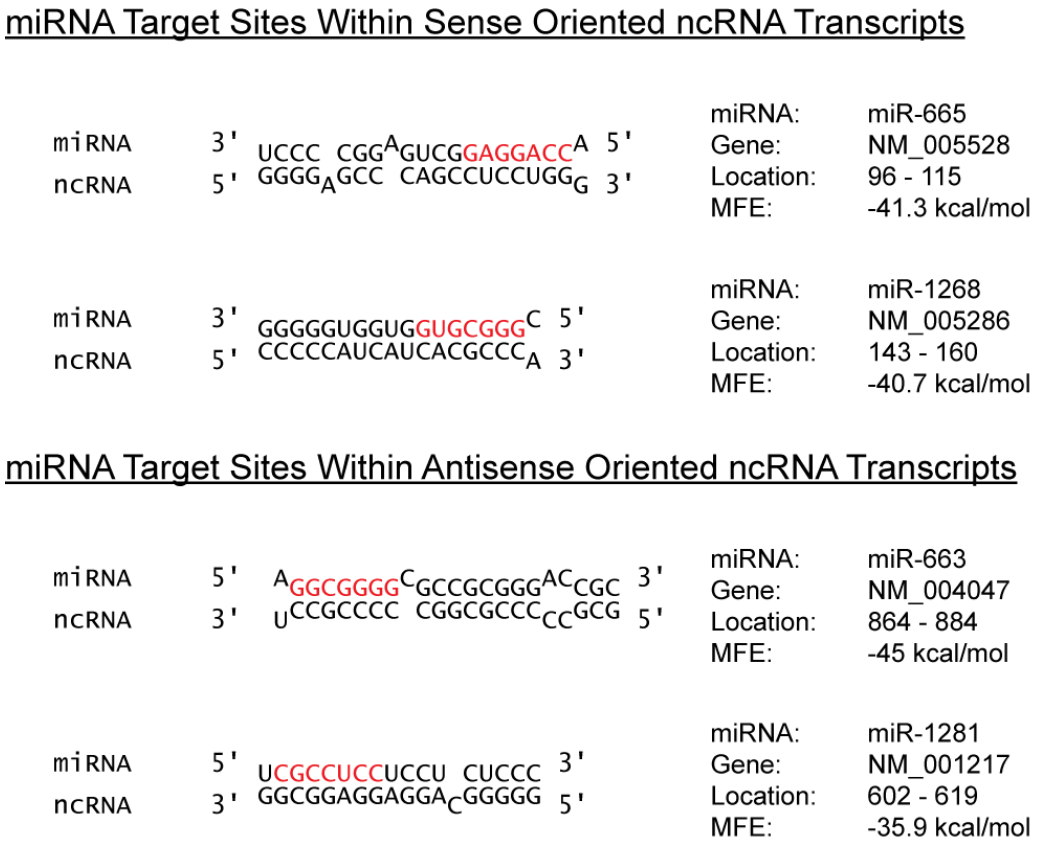
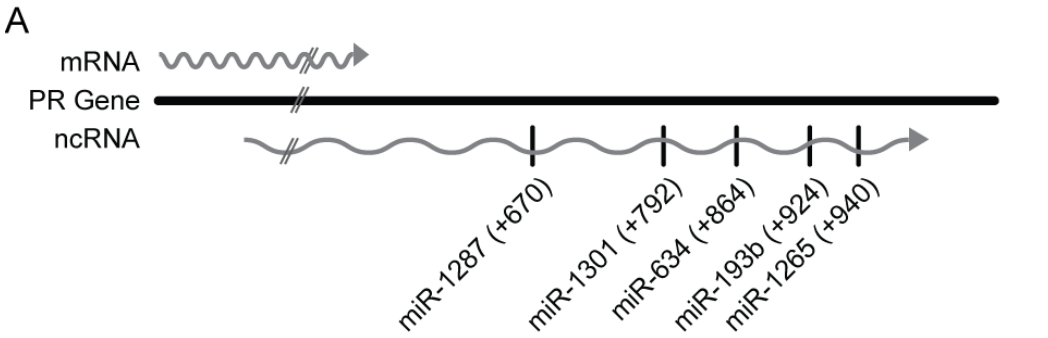


Figure 4.4. miRNA target sites within the ncRNA that overlaps the PR gene terminus. (A) Schematic of selected miRNA target sites within the target ncRNA. (B) Alignments of selected miRNAs with the target ncRNA overlapping the PR gene terminus (seed sequences shown in red).



B

```
ncRNA       5' UAUUGGAGCAACAGCAUCCAGC 3'
miR-1287    3' CUGAGCUUGGUGAC UAGGUCGU 5'

ncRNA       5'    UUCACAGAUUUUCUGGAAAAGCUGCA 3'
miR-1301    3' CUUCAGUG    AGGGUCC G UCGACGUU 5'

ncRNA       5' UUGUAAAGUGUCAUGGUGCUGGU 3'
miR-634     3' CAGGUUUCA ACC CCACGACCAA 5'

ncRNA       5' UAGCU GAUUU GGGCCAGU 3'
miR-193b    3'  UCGCCCUGAAACUCCCGGUCAA 5'

ncRNA       5' C CAGUUUCUUAGCUACAUCCU 3'
miR-1265    3' UUGUUGU GAACUGGUGUAGGAC 5'
```

Figure 4.5. miRNAs complementary to the PR terminus inhibit PR protein expression. Western analysis showing inhibition of PR protein expression by miRNA mimics. miRNA mimics were added to cells at 25 nM.

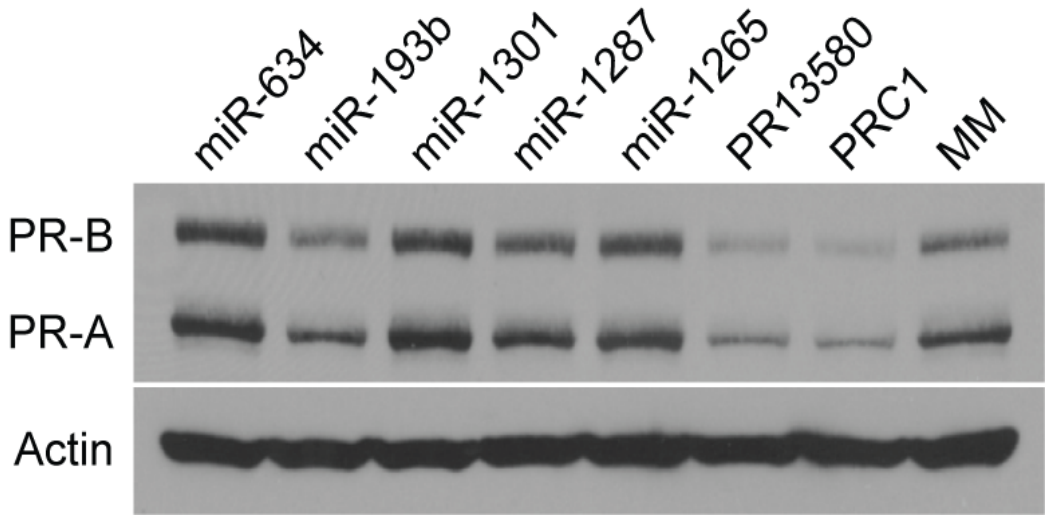


Figure 4.6. Dose dependent inhibition of PR by miR-193b. (A) Western analysis showing inhibition of PR protein expression by miR-193b. (B) Quantification of independent dose response experiments measuring reduction of PR protein levels by miR-193b. Error bars indicate s.d. (n = 4).

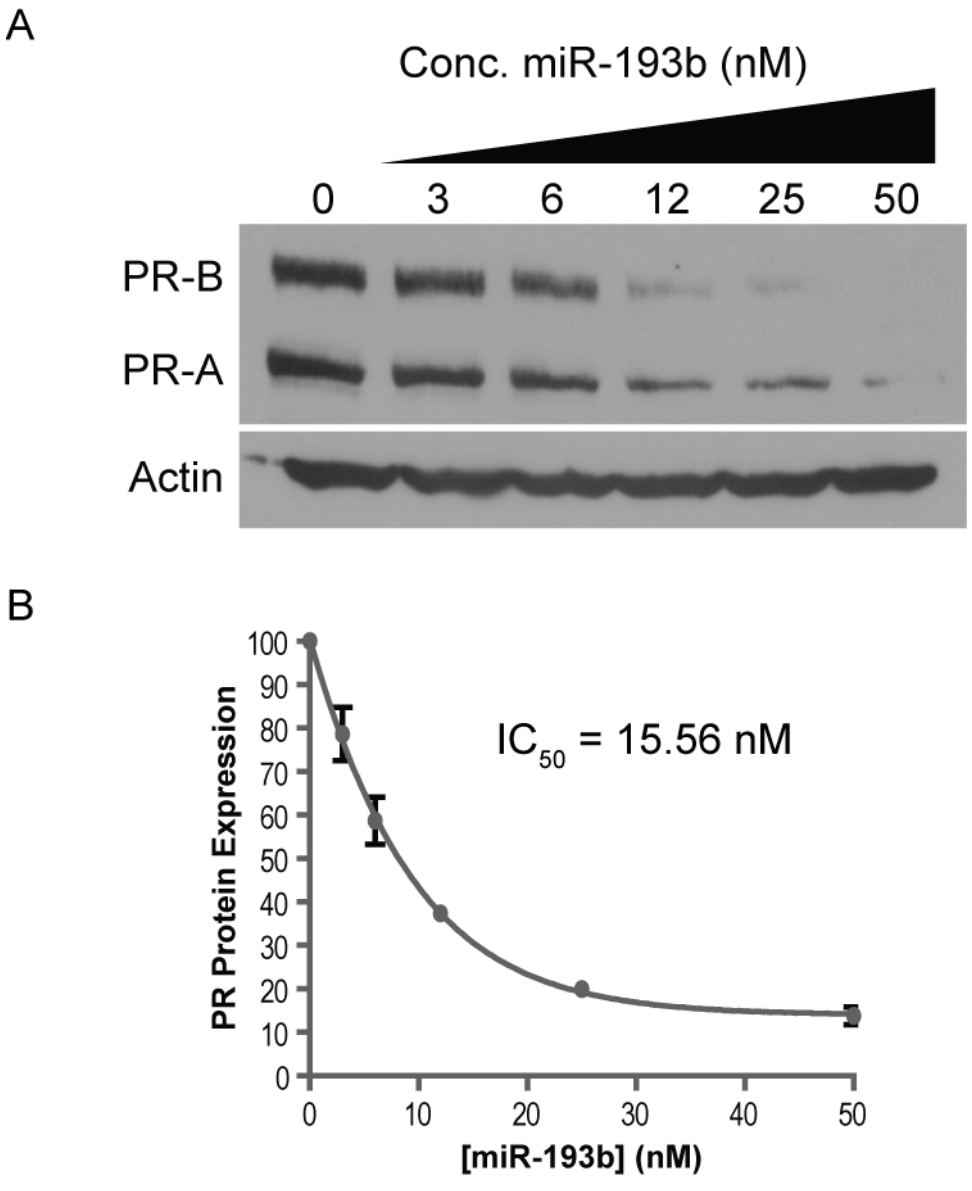


Figure 4.7. miR-193b inhibits PR RNA expression. (A) RT-qPCR measuring PR mRNA expression. (B). RT-qPCR measuring expression of the ncRNA overlapping the PR gene terminus. miRNA mimics were added to cells at 25 nM. Error bars indicate s.d. (n = 3). *P*-values were calculated using the two-tailed unpaired Student's *t*-test with equal variances. ** $P < 0.01$.

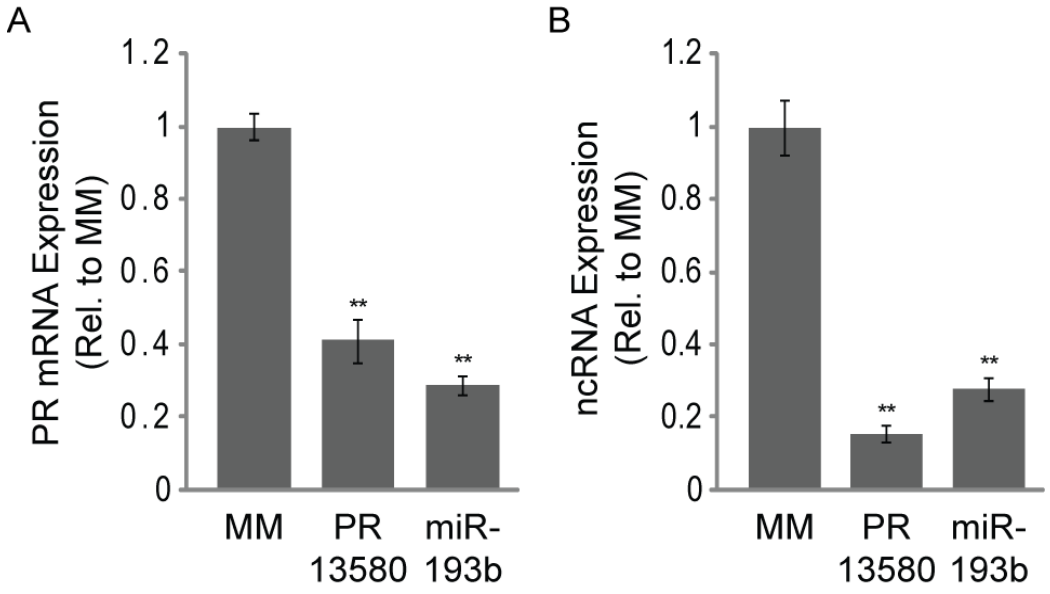


Figure 4.8. miR-193b inhibits PR transcription. (A) RT-qPCR measuring PR hnRNA expression. (B) ChIP-qPCR measuring Pol II occupancy on the PR promoter. miRNA mimics and siRNAs were added to cells at 25nM. Error bars indicate s.d. (n=3). *P*-values were calculated using the two-tailed unpaired Student's t-test with equal variances. * $P < 0.05$, ** $P < 0.01$.

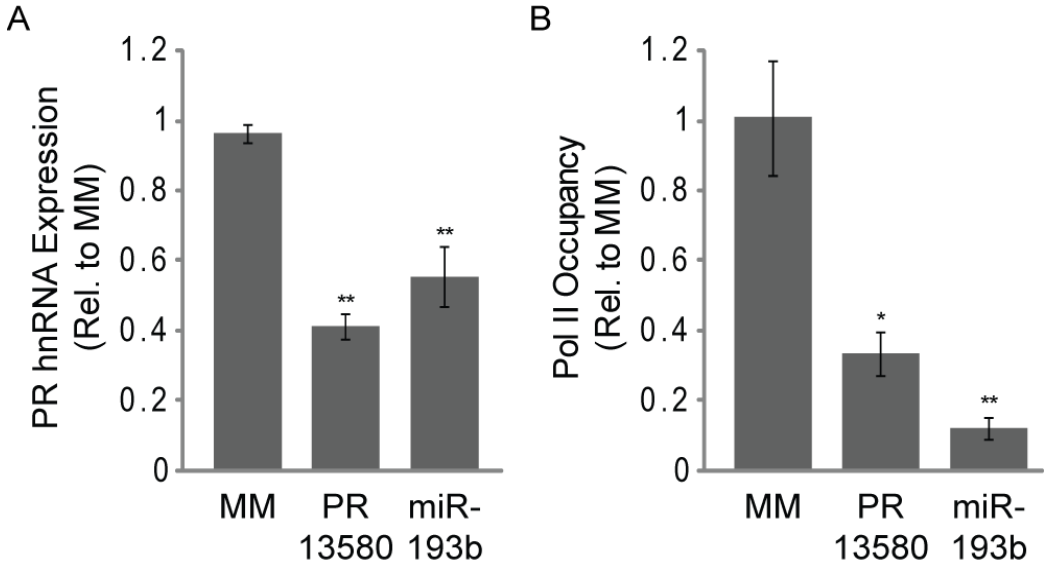
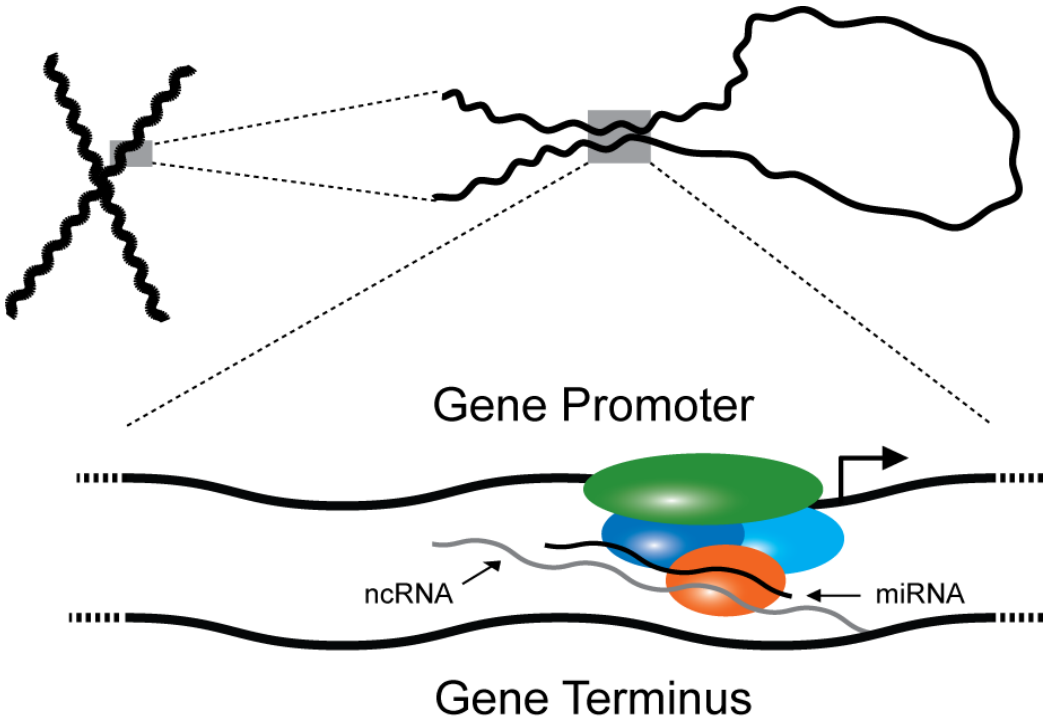


Figure 4.9. Model for mechanism.



References For Chapter 4

1. Gingeras, T. R. Origin of phenotypes: genes and transcripts. *Genome Res* **17**, 682-690 (2007).
2. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-1488 (2007).
3. Schwartz, J. C. *et al.* Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol* **15**, 842-848 (2008).
4. Yue, X. *et al.* Transcriptional regulation by small RNAs at sequences downstream from 3' gene termini. *Nat Chem Biol* **6**, 621-629 (2010).
5. Chu, Y., Yue, X., Younger, S. T., Janowski, B. A. & Corey, D. R. Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. *Nucleic Acids Res* **38**, 7736-7748 (2010).
6. Janowski, B. A. *et al.* Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat Chem Biol* **1**, 216-222 (2005).
7. Janowski, B. A. *et al.* Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* **13**, 787-792 (2006).
8. Janowski, B. A. *et al.* Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol* **3**, 166-173 (2007).
9. Ambros, V. The evolution of our thinking about microRNAs. *Nat Med* **14**, 1036-1040 (2008).
10. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233 (2009).
11. Kim, D. H., Saetrom, P., Snove, O., Jr. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 16230-16235 (2008).
12. Majid, S. *et al.* MicroRNA-205-directed transcriptional activation of tumor suppressor genes in prostate cancer. *Cancer* **116**, 5637-5649 (2010).
13. Place, R. F., Li, L. C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* **105**, 1608-1613 (2008).
14. Younger, S. T., Pertsemlidis, A. & Corey, D. R. Predicting potential miRNA target sites within gene promoters. *Bioorg Med Chem Lett* **19**, 3791-3794 (2009).
15. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443-453 (1970).
16. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-1050 (2005).
17. Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput Biol* **2**, e33 (2006).
18. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-144 (2006).
19. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840 (2010).

20. Brown, C. J. *et al.* 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**, 527-542 (1992).
21. Hirota, K. *et al.* Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* **456**, 130-134 (2008).
22. Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717-1720 (2008).
23. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137 (1996).
24. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007).
25. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**, 126-130 (2008).
26. Wutz, A. *et al.* Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745-749 (1997).
27. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).
28. Wahlestedt, C. Natural antisense and noncoding RNA transcripts as potential drug targets. *Drug Discov Today* **11**, 503-508 (2006).
29. Calin, G. A. *et al.* Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **12**, 215-229 (2007).
30. O'Sullivan, J. M. *et al.* Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* **36**, 1014-1018 (2004).
31. Tan-Wong, S. M., French, J. D., Proudfoot, N. J. & Brown, M. A. Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proc Natl Acad Sci U S A* **105**, 5160-5165 (2008).
32. Tiwari, V. K. *et al.* PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol* **6**, 2911-2927 (2008).
33. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

Chapter 5: Methods

5.1 Cell culture

T47D and MCF7 breast cancer cells (American Type Culture Collection, ATCC) were maintained in RPMI-1640 medium (ATCC) supplemented with 10% (v/v) FBS (Atlanta Biologicals), 0.5% (v/v) nonessential amino acids (Sigma), 10mM HEPES (Sigma), 1mM Sodium Pyruvate (Sigma), 0.4 units ml^{-1} bovine insulin (Sigma). Cells were cultured at 37 °C and 5% CO_2.

5.2 Cellular delivery of miRNA mimics and siRNAs

RNAiMAX (Invitrogen) was used to deliver small duplex RNAs into T47D or MCF7 cells as per the manufacturer's instructions. For RNA and protein isolation, cells were plated in 6-well dishes at densities ranging between 150K-200K cells/well. For RNA and chromatin immunoprecipitation, cells were plated in 10 cm^2 dishes at a density of $4.5x10^6$ cells/dish. Cells were transfected 48 hours after plating. A double transfection protocol was used for AGO reversal experiments. The first transfection was performed as described above (using mismatch-containing duplexes or siRNAs against AGO). After 72 hours, cells were dissociated using trypsin and re-seeded in transfection reagent (using MM, miR-423-5p, or pre-miR-423 duplexes). Sequences for miRNA mimics and siRNAs are listed in Figure 5.1.

5.3 Western blotting

Cells were harvested 5 days post-transfection for protein isolation. Cell pellets were lysed and protein concentrations were quantified by BCA assay (Pierce). Western blots

were performed on protein lysates (30 μg per well). Primary antibodies used were α-PR (Cell Signaling Technology) and α-β-actin (Sigma). Protein was visualized with horseradish peroxidase-conjugated α-mouse secondary antibody (Jackson Immunolabs) and Supersignal developing solution (Pierce).

5.4 Quantitative PCR

Cells were harvested 72 hours post-transfection for RNA isolation. RNA was isolated using TRI Reagent (Sigma) as per the manufacturer's instructions. For each sample, 2 μg of RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). RNA was treated with DNase I (Worthington) prior to reverse transcription. qPCR was performed on an ABI7900 real-time PCR (Applied Biosystems) using iTaq SYBR Green Supermix (Bio-Rad). ABI TaqMan miRNA assays were used to detect miR-423-5p and U6 snRNA as per the manufacturer's instructions. Primers for PR and GAPDH mRNA were supplied by Applied Biosystems. All additional primers were designed using Primer3. Only those primer sets that showed linear amplification over several orders of magnitude were used for quantification. Primers and PCR conditions are listed in Supplementary Figure 5.2.

5.5 Chromatin immunoprecipitation

Cells were harvested 72 hours post-transfection for chromatin isolation. Dishes were incubated in 1% formaldehyde in PBS for 10 min followed by addition of glycine to a final concentration of 0.125 M to quench the crosslinking reaction. Cells were then washed with PBS and transferred to 15 ml conical tubes. Nuclear fractions were isolated

using two successive washes with hypotonic lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and 0.5% Nonidet P40). Nuclei were lysed in 1 ml of nuclear lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0, and 1X complete protease inhibitor cocktail (Roche)). DNA was sheared by sonication and insoluble material was removed by centrifugation. For each experiment 10 ul of sample was removed for input normalization. For each IP 100 ul of sample was diluted to 1 ml with IP buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl, and 1X complete protease inhibitor cocktail (Roche)). Diluted samples were incubated with 4 ug of antibody overnight at 4°C and chromatin:antibody complexes were captured using Protein G Plus/Protein A agarose beads (Calbiochem). α-RNA Polymerase II, α-H3K9me2, normal rabbit IgG, and normal mouse IgG antibodies were supplied by Millipore. Beads were washed with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, and 150 mM NaCl), high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, and 500 mM NaCl), LiCl wash buffer (0.25 M LiCl, 1% Nonidet P40, 1% deoxycholate, 1 mM EDTA, and 10 mM Tris-HCl pH 8.0), and two washes with Tris-EDTA pH 8.0. Chromatin complexes were eluted from beads using two successive 15 min incubations in 250 ul of elution buffer (1% SDS and 0.1 M $NaHCO_3$). Eluates were combined, NaCl was added to each sample to a final concentration of 200 nM, and samples were incubated overnight at 65°C to reverse crosslinks. Samples were treated with 1 ul of RNase A and incubated at 37°C for 30 min followed by addition of 20 ug of proteinase K, 20 ul of Tris-HCl pH 7.0, 10 ul of 0.5 M EDTA and incubation at 42°C for 45 min. DNA was isolated using phenol-chloroform extraction followed by isopropanol

precipitation. qPCR analysis of isolated DNA was performed on an ABI7900 real-time PCR (Applied Biosystems) using iTaq SYBR Green Supermix (Bio-Rad). Primers for the PR promoter and GAPDH promoter were designed using Primer3. Only those primer sets that showed linear amplification over several orders of magnitude were used for quantification. Primers and PCR conditions are listed in Figure 5.2.

5.6 RNA immunoprecipitation

Cells were harvested 72 hours post-transfection for RNA isolation. Dishes were incubated in 1% formaldehyde in PBS for 10 min followed by addition of glycine to a final concentration of 0.125 M to quench the crosslinking reaction. Cells were then washed with PBS and transferred to 15 ml conical tubes. Nuclear fractions were isolated using two successive washes with hypotonic lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and 0.5% Nonidet P40). Nuclei were lysed in 1 ml of nuclear lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0, 1X complete protease inhibitor cocktail (Roche), and 50 U/ml RNase inhibitor (Promega)). For each experiment 10 ul of sample was removed for input normalization. For each IP 100 ul of sample was diluted to 1 ml with IP buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl, 1X complete protease inhibitor cocktail (Roche), and 50 U/ml RNase inhibitor (Promega)). Diluted samples were incubated with 4 ug of antibody overnight at 4°C and RNA:protein:antibody complexes were captured using Protein G Plus/Protein A agarose beads (Calbiochem). α-AGO1, α-AGO2, normal rabbit IgG, and normal mouse IgG antibodies were supplied by Millipore. Beads were washed with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2

mM EDTA, 20 mM Tris-HCl pH 8.0, and 150 mM NaCl), high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, and 500 mM NaCl), LiCl wash buffer (0.25 M LiCl, 1% Nonidet P40, 1% deoxycholate, 1 mM EDTA, and 10 mM Tris-HCl pH 8.0), and two washes with Tris-EDTA pH 8.0. RNA:protein complexes were eluted from beads using two successive 15 min incubations in 250 ul of elution buffer (1% SDS, 0.1 M $NaHCO_3$, and 50 U/ml RNase inhibitor (Promega)). Eluates were combined, NaCl was added to each sample to a final concentration of 200 nM, and samples were incubated for 4 h at 65°C to reverse crosslinks. Samples were treated with 20 ug of proteinase K, 20 ul of Tris-HCl pH 7.0, 10 ul of 0.5 M EDTA and incubated at 42°C for 45 min. RNA was isolated using phenol-chloroform extraction followed by isopropanol precipitation. RNA was treated with DNase I (Worthington) and reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). PCR amplification of cDNA was performed using HotStarTaq DNA polymerase (QIAGEN) as per the manufacturer's instructions. Primers for the PR promoter were designed using Primer3. Primers and PCR conditions are listed in Figure 5.2. PCR products were resolved on 3% agarose gel with ethidium bromide.

5.7 Methylation specific PCR and bisulfite sequencing

Genomic DNA was extracted using TRI Reagent (Sigma) as per the manufacturer's instructions. For each sample, 2 µg of DNA was diluted into 50 µl of H_2O. Single stranded DNA was created by adding 5.5 µl of 2 M NaOH and incubating at 37°C for 10 min. To each sample, 30 µl of freshly prepared 10 mM hydroquinone (Sigma) and 520 µl of freshly prepared 3 M sodium bisulfite (Sigma) were added. DNA was incubated at

50°C for 16 h. After bisulfite treatment, DNA was purified using the Wizard DNA Clean-Up System (Promega). DNA was then subjected to ammonium acetate precipitation and ethanol washing to remove any remaining impurities. PCR amplification of treated DNA was performed using HotStarTaq DNA polymerase (QIAGEN) as per the manufacturer's instructions. Primers and PCR conditions are listed in Figure 5.2. PCR products were resolved on 3% agarose gel with ethidium bromide. For bisulfite sequencing PCR products were cloned into a PCR-4 Topo vector (Invitrogen) and sequenced (McDermott sequencing core, University of Texas Southwestern).

5.8 Statistical Analysis

All data are presented as means ± standard deviation of three or more independent results. Statistical significance was assessed using a two-tailed unpaired Student's *t*-test.

Figure 5.1. Sequences of siRNAs and miRNA mimics. All RNAs were transfected as duplexes consisting of the sequence listed and a fully complementary RNA strand unless noted otherwise in the text. *** indicates DNA oligonucleotides, ¶ indicates 2'-O-methyl RNA oligonucleotides. † siRNAs targeting AGO2 were introduced as a pool of multiple siRNAs.

Oligo Name	Target	Sequence
miR-548k	PR Promoter	AAAAGUACUUGCGGAUUUUGCU
miR-1278	PR Promoter	UAGUACUGUGCAUAUCAUCUAU
miR-372	PR Promoter	AAAGUGCUGCGACAUUUGAGCGU
miR-373	PR Promoter	GAAGUGCUUCGAUUUUGGGGUGU
miR-520c-3p	PR Promoter	AAAGUGCUUCCUUUUAGAGGGU
miR-595	PR Promoter	GAAGUGUGCCGUGGUGUGUCU
miR-532-3p	PR Promoter	CCUCCCACACCCAAGGCUUGCA
miR-214	PR Promoter	ACAGCAGGCACAGACAGGCAGU
miR-525-5p	PR Promoter	CUCCAGAGGGAUGCACUUUCU
miR-423-5p	PR Promoter	UGAGGGGCAGAGAGCGAGACUUU
miR-423-3p	N/A	AGCUCGGUCUGAGGCCCCUCAGU
PR9	PR Promoter	UGUCUGGCCAGUCCACAGC dTdT
miR-634	PR Terminus	AACCAGCACCCCAACUUUGGAC
miR-193b	PR Terminus	AACUGGCCCUCAAAGUCCCGCU
miR-1301	PR Terminus	UUGCAGCUGCCUGGGAGUGACUUC
miR-1287	PR Terminus	UGCUGGAUCAGUGGUUCGAGUC
miR-1265	PR Terminus	CAGGAUGUGGUCAAGUGUUGUU
PR13580	PR Terminus	GUUUGCCUGCAUCAGUUCC dTdT
PRC1	PR mRNA	AUGGAAGGGCAGCACAACU dTdT
MM	N/A	UCUCUCGCGAGUCCACAGC dTdT
Seed-MM	PR Promoter	UGAAAAGCAGAGAGCGAGACUUU
Mid-MM	PR Promoter	UGAGGGGCAGAGACGCAGACUUU
End-MM	PR Promoter	UGAGGGGCAGAGAGCGAGAGAAU
MM-DNA***	N/A	TCTCTCGCGAGTCCACAGC dTdT
miR-423-5p-DNA***	PR Promoter	TGAGGGGCAGAGAGCGAGACTTT
siAgo1	AGO1	GGAGUUACUUUCAUAGCAUUU
siAgo2-1†	AGO2	GCACGGAAGUCCAUCUGAAUU
siAgo2-2†	AGO2	GCAGGACAAAGAUGUAUUAUU
siAgo2-3†	AGO2	GGGUCUGUGGUGAUAAAUAUU
siAgo2-4†	AGO2	GUAUGAGAACCCAAUGUCAUU
2'-anti-miR¶	miR-423-5p	AAAGUCUCGCUCUCUGCCCCUCA
2'-MM¶	N/A	AAUGUCACGCACUCAGCCGCUCU
2'-SCR¶	N/A	CCCUCACUCUGCCUCGCUAAAGU

Figure 5.2. PCR primer sequences and conditions. ‡ Touchdown 65-55°C, 45 s over 10 cycles followed by 30 cycles at 55°C, 45 s.

Target	Direction	Use	Sequence
PR hnRNA	Forward	qPCR	TCCTCACTAGCTCCCCTTCA
PR hnRNA	Reverse	qPCR	TGATTGGATTTTGGGTGGTT
PR promoter	Forward	ChIP/RIP	CCTAGAGGAGGAGGCGTTGT
PR promoter	Reverse	ChIP/RIP	ATTGAGAATGCCACCCACA
GAPDH promoter	Forward	ChIP	TACTAGCGGTTTTACGGGCG
GAPDH promoter	Reverse	ChIP	TCGAACAGGAGGAGCAGAGAGCGA
unmethylated PR promoter	Forward	MSP	TGATTGTTGTTTGTAGTATG
unmethylated PR promoter	Reverse	MSP	CAACAATTTAATAACACACA
methylated PR promoter	Forward	MSP	TGATTGTCGTTCGTAGTACG
methylated PR promoter	Reverse	MSP	CGACAATTTAATAACACGCG
PR promoter	Forward	Sequencing	GTTTTATTTTAAAGAATTTGTTATTGAGAG
PR promoter	Reverse	Sequencing	AATCAACTCCTACCCTTAACCTCCAT
IGSF1 mRNA	Forward	qPCR	GAATTGGCCACATTTGCTTT
IGSF1 mRNA	Reverse	qPCR	AGGAAGCCAAAGCACAGAAA
IGSF1 hnRNA	Forward	qPCR	CTGGAGGAGCTCACTGGAGA
IGSF1 hnRNA	Reverse	qPCR	GGGAGCAGTTTGATTTACGG
IGSF1 promoter	Forward	RT-PCR	CTGAATTTGGCTCCAGCAAC
IGSF1 promoter	Reverse	RT-PCR	GAGCCCACCTACACCAACTC

PCR Conditions	Denature	Annealing	Extension
RT-qPCR	95°C, 20 s	57°C, 30 s	72°C, 45 s
ChIP-qPCR	95°C, 20 s	57°C, 30 s	72°C, 45 s
MSP-PR-Unmethylated	94°C, 30 s	45°C, 30 s	72°C, 60 s
MSP-PR-Methylated	94°C, 30 s	51°C, 45 s	72°C, 60 s
Bisulfite Sequencing	94°C, 50 s	65°C, 45 s‡	72°C, 60 s

Chapter 6: Future directions and concluding remarks

6.1 Introduction

The work within this dissertation demonstrates that miRNAs are capable of regulating gene transcription through recognition of ncRNA transcripts that overlap gene promoters. In addition, this work suggests that miRNAs are capable of regulating transcription by targeting sequences downstream of the 3'-terminus of protein-coding genes. The experiments presented in this dissertation have relied on the use of exogenously added miRNA mimics to characterize the mechanism of transcriptional silencing and future experiments will be needed to demonstrate that endogenous miRNAs perform this function.

6.2 Identifying endogenous miRNAs that regulate transcription

Experiments aimed at the identification of endogenous miRNAs that regulate transcription are more complicated than experiments that use miRNA mimics to study mechanism. miRNA target prediction algorithms can be helpful for suggesting interactions between miRNAs and their putative target sequences, but validation requires a model system in which both the miRNA and the target RNA are expressed. For traditional miRNA interactions with mRNA, gene expression microarrays have been successful at validating miRNA targets by complementing computational target predictions with anti-correlation data between miRNA and mRNA transcript expression. However, the mechanism for miRNA-induced transcriptional silencing requires the expression of ncRNA transcripts that overlap either gene promoters or gene termini. This

additional layer of complexity reduces the effectiveness of simply monitoring miRNA and mRNA expression.

Recent advances in RNA sequencing technology provide new approaches for studying RNA-protein interactions that were not available at the inception of this work. It is now possible to isolate a protein of interest and sequence all associated nucleic acid in a cost-effective manner. Figures 3.15 and 3.17 of this dissertation demonstrate that miRNA-induced transcriptional silencing is mediated by AGO proteins, specifically AGO2. In an experiment similar to Figure 3.15B, AGO2 and associated RNA transcripts will be immunoprecipitated from nuclear fractions of T47D breast cancer cells. As opposed to validating AGO2 association with RNA using RT-PCR, isolated RNA will be subjected to deep sequencing. This approach will provide an unbiased, genome-wide analysis of AGO2 associated RNA transcripts.

Sequencing reads will be mapped back to their genomic origins and special emphasis will be placed on reads that map to promoter regions of protein-coding genes. The computational approaches described in chapter 2 of this dissertation will be used to further select for isolated transcripts that have promising predicted miRNA target sites. As a final filter, nuclear miRNA expression will be assessed using miRNA microarrays and AGO2 associated transcripts that contain predicted miRNA target sites for expressed miRNAs will be selected for experimental validation.

Identified transcripts that meet the previously described criteria will be validated using RT-PCR with RNA isolated from nuclear fractions. Once validated, RIP followed by RT-PCR will be used to confirm transcript association with AGO2. To determine if AGO2 associated transcripts are involved in transcriptional regulation by miRNAs,

inhibitors will be designed against miRNAs with strong complementarity to the transcripts. Inhibitors will consist of fully modified 2'-O-methyl RNAs that are fully complementary to their target miRNAs. Inhibitors will be introduced into T47D cells and their effect on gene expression will be assessed using RT-qPCR.

If the inhibition of miRNAs that are predicted to target AGO2-associated transcripts that overlap gene promoters has a significant effect of the expression of downstream genes, the mechanism of regulation will be further characterized. Specifically, the effect of miRNA inhibitors on gene transcription will be monitored by chromatin immunoprecipitation (ChIP) using antibodies against RNA Polymerase II (Pol II). Furthermore, the effect of inhibitor addition on chromatin structure will be assessed by ChIP using antibodies against a variety of histone modifications.

The experimental approach outlined here has been focused on AGO2-associated transcripts that overlap gene promoters. This same experimental workflow can be applied to additional classes of transcripts identified using RNA-seq. For example, AGO2-associated transcripts that overlap gene termini are also of significant interest.

6.3 Identifying proteins involved in miRNA-induced transcriptional silencing

To date, the only proteins implicated in the mechanism of small RNA-induced transcriptional regulation are the AGOs. However, recent advances in high-throughput proteomic approaches have made it possible to isolate a protein of interest and identify associated proteins in an unbiased manner. In experiments similar to those described in section 6.2, AGO2 will be immunopurified from nuclear fractions of T47D breast cancer

cells. Protein isolates will be subjected to 2-dimensional nano-liquid chromatography tandem mass spectrometry (2D-LC-MS/MS).

Proteins identified using 2D-LC-MS/MS will be ranked based on their abundance within the analyzed samples. Highly abundant proteins will be further prioritized based on their function. For example, it has been shown that transcriptional regulation by small RNAs is associated with changes in histone modifications and emphasis will be placed on validating the association of AGO2 with known chromatin modifying enzymes. For validation, AGO2 will be immunopurified from nuclear fractions of T47D cells and protein isolates will be probed with antibodies against proteins of interest. In addition, reciprocal co-immunoprecipitations will be formed where a protein of interest is immunopurified from nuclear fractions and protein isolates will be probed with antibodies against AGO2.

To further characterize the role of identified proteins in small RNA-induced transcriptional silencing, siRNA knockdown experiments will be used. siRNAs will be designed against mRNAs that encode proteins associated with AGO2. Following siRNA knockdown, the expression of genes that are transcriptionally regulated by miRNAs (identified in section 6.2) will be monitored using RT-qPCR. Significant emphasis will be placed on genes in which their expression is required for transcriptional regulation by small RNAs.

6.4 Concluding remarks

Completion of the experiments outlined in this chapter will further establish a role for miRNAs in transcriptional regulation. In addition, the unbiased nature of these

approaches has the potential to identify novel functions of miRNAs that have been previously unappreciated.