SMALL RNAS REGULATE TRANSCRIPTION BY INTERACTING WITH NONCODING RNA TRANSCRIPTS

APPROVED BY SUPERVISORY COMMITTEE

David R. Corey, Ph.D., Advisor

Hongtao Yu, Ph.D., Chairman

Rolf A. Brekken, Ph.D.

Kevin Gardner, Ph.D.

DEDICATION

To my dearest friend, Emese Dian, and to Caleb Schwartz, my son.

SMALL RNAS REGULATE TRANSCRIPTION BY INTERACTING WITH NONCODING RNA TRANSCRIPTS

by

JACOB C. SCHWARTZ

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

December, 2009

ACKNOWLEDGEMENTS

I did not start in David Corey's lab with the intention of pursuing doctoral work. I had presumed I already knew what a professional scientist did. However, David patiently exposed me to a side of professional science that I had not experienced. In so doing, David helped me realize that this profession is truly my calling. Without David's coaching, I would have given up on this career before I even began.

David not only got me started but he spent extraordinary time and effort in training me with the basic skill sets that scientists need to be successful – grant writing, paper writing, speaking, experimental design, and keeping current on scientific literature. I cannot thank him enough for his devotion to my training, the same devotion that David offers to all of his students as he puts every effort in helping them achieve their full potential.

I am also grateful to Bethany Janowski, who allowed me to work alongside her as she pioneered antigene technologies. She mentored me in techniques such as mammalian tissue culture and constantly helped me keep my experiments running. We worked side-by-side on every project that I undertook in the Corey lab and her undaunted drive for breaking into new experimental territory and paving new paths has left an indelible impression on how I direct and think about my research.

In the Corey lab, my every day was enriched by the training and companionship I gained from my labmates. From the people in the lab when I first joined, including Kenneth Huffman, Zain Paroo, Kunihiro Kaihatsu, Janet Freisen, and Alan Meschek, to those that joined soon after including Jiaxin Hu, Rosie Ram, Randall Beane, Becky Nguyen, and Scott Younger, and those that came along the way including Yongjun Chu, Adele Yue, Masayuki Matsui, and Amritha Bhat, these people made me look forward with excitement to each day that I came to the lab. Toward the end of my tenure in the Corey lab, my experience was enriched by the addition of Jonathan Watts and Keith Gagnon to the lab, who offered me guidance from their own experience as I began the transition from a graduate career to embark on a post-graduate career. It has been a great priveledge to work with these incredibly good natured and gifted scientists.

The faculty at UT Southwestern has been universally supportive and nuturing towards me at every step in my graduate career. I doubt that there is a more nurturing doctoral training environment with professors more devoted towards student enrichment than at UT Southwestern. A long list of professors in the departments of Pharmacology and Biochemistry has taken time to help me and teach me through my graduate career and I am very grateful to them.

I have had the pleasure of having a fantastic thesis committee. Each of my members, Rolf Brekken, Hongtao Yu, and Kevin Gardner have put in extra efforts to support me at every opportunity they have been given. Their letters of support and recommendation have paved the way for me to achieve more than I would have ever dared to hope for during my graduate career.

While at UT Southwestern, I have enjoyed enriching collaborations with other labs at UT Southwestern, including the Brugarolas and Minna labs. I have also enjoyed collaborations with Daniel Hardy in Carol Mendelsonn's lab, Bret Monia from Isis Pharmaceutics, and Sayda Elbashir at Alnylam Pharmaceutics.

I received daily support and encouragement from my fiancé Emese Dian and my pets: Shiba, Junior, Aura, Seabass, Buddy, Trypsin, Watson, Crick, Csilla, Norbert, and Mr. Frog. If it was not for their love and support, I would not have the energy to do what I do. And Emese's family has buttressed her support for me and I am grateful for them. My son, Caleb, has been a delight and motivation for me to try to push myself to achieve more than I thought I could.

Along my academic development, I received essential encouragement and guidance from former professors whose labs I worked in, including Daniel Marble, Sam Matteson, Zhibing Hu, Jerry Duggan, Floyd McDaniel, and Gunter Gross. I have been impressed countless times how the lessons I learned under their instruction have a dialy impact my research and career. They have truly had a lasting impact on my scientific development.

Last but certainly not least, I thank my family. My Mom and Dad have endowed me with both the nature and nurturing to achieve all that I have and will achieve. I am not only grateful for their nurturing during my development but also their constant support

and encouragement I enjoy day by day. Their support is also buttressed by my step-dad, Tommy Moczygemba, and my step-mom, Becky Powell. My life has been greatly enriched by the addition of them to my family. I thank my sister Sarah and her husband Matt Horner for their companionship. My extended family from both my Mom's and Dad's sides has constantly peppered my graduate career with their praise and encouragement and I very much appreciate them.

SMALL RNAS REGULATE TRANSCRIPTION BY INTERACTING WITH NONCODING RNA TRANSCRIPTS

Jacob C. Schwartz, M.S.

The University of Texas Southwestern Medical Center at Dallas, 2009

Supervising Professor: David R. Corey, Ph.D.

General methods for controlling gene expression have long been appreciated as an attractive target in drug design. Recently, the Corey lab has demonstrated that short RNA duplexes designed to target the promoter region for human genes can inhibit or activate gene expression in a sequence dependent manner. The mechanism by which RNAs achieve promoter recognition has remained unclear. Sequence specific recognition could be achieved by (1) RNA hybridization to genomic DNA, or (2) RNA recognition of some uncharacterized RNA species. Promoter targeted duplex RNA has been shown to recruit argonaute proteins to the promoter DNA and these proteins are necessary for duplex RNAs to regulate transcription. Argonaute proteins are known to recognize RNA:RNA interactions. However, genes targeted with duplex RNAs have no characterized transcripts in their promoters. I tested the hypothesis that promoter RNA transcripts exist and serve as a substrate for short duplex RNAs to hybridize to and regulate gene expression of adjacent genes.

I found previously undiscovered RNA transcripts expressed from the promoter of progesterone receptor (PR) using RT-PCR. Quantitative RT-PCR of the promoter RNA of PR reveals expression levels between 10 and 1000 fold lower than PR in T47D and MCF7 breast cancer cells. I have cloned three transcripts overlapping the promoter of PR from two cell lines – T47D and MCF7, each with unique splicing and transcription start sites. All of these transcripts initiate within the protein coding region of PR and run antisense to the gene PR. I have been able to show that the promoter transcripts can be immunoprecipitated with antibodies against the argonaute proteins in cells transfected

with duplex RNAs targeting the promoter of PR but not in cells transfected with mismatched duplex RNAs. Also, biotinylated RNAs bind to and pull down these noncoding RNAs. Finally, knockdown of the antisense transcript with an antisense oligonucleotide prevent gene activation by duplex RNAs.

Following this study, our lab uncovered that duplex RNAs can target beyond the 3' terminus of genes and silence or activate transcription. I further showed that this transcription regulation is mediated by argonaute binding to noncoding RNAs overlapping the 3' terminus of the genes, PR and BRCA1. The signal is transmitted from the 3' terminus to the gene promoter because the 5' and 3' ends of these genes are held in a chromatin loop, which I validated using a chromatin conformation capture assay. This brings the ends of the gene in close proximity to each other. Due to this interaction, short RNAs that bind a noncoding RNA at the 3' end of the gene also physically interacts with noncoding RNAs that associate with the gene promoter. This is confirmed by RNA immunoprecipitation of both transcripts with duplex RNAs targeting either the 5' or 3' ends of the gene.

More than 20 years ago, it was found that proteins recognizing DNA at the 5’ end of genes could regulate transcription. This study presents a paradigm shift implicating noncoding RNAs at the 5' and 3' ends of genes can be recognized by proteins which activate or inhibit transcription of adjacent protein coding genes. Recent studies demonstrate an abundance of RNAs transcribed in human cells that do not code for protein. My results suggest a new model for duplex RNA recognition of gene promoters. Argonaute proteins loaded with one strand of the RNA duplex recognizes, through Watson-Crick base pairing, a noncoding transcript that is associated with chromatin at the promoter of the targeted gene. This RNA:RNA interaction in close proximity to the promoter mediates protein-protein interactions between argonaute and other factors on the promoter to turn off or on gene expression.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

1. Yue A, **Schwartz JC**, Younger ST, Chu Y, Gagnon KT, Elbashir S, Janowski BA, Corey DR, "Regulation of Transcription by Small RNAs Complementary to Sequences Downstream from the 3' Termini of Genes", *Cell*, 2009 (*In review*)
2. Hu J, Matsui M, Gagnon KT, **Schwartz JC**, Gabillet S, Arar K, Wu J, Bezprozvanny I, Corey DR, "Inhibiting Expression of Mutant Huntingtin and Ataxin-3 by Targeting Expanded CAG Repeat RNAs", *Nat Biotech*, 2009 (*In press*)
3. **Schwartz, JC**, Younger, S, Nguyen,N, Hardy, D, Monia, BP, Corey, DR, Janowski, BA, "Antisense Transcripts are Targets for Small Activating RNAs", *Nat Struct Mol Biol.* 2008 Aug;15(8):842-8.
4. Janowski BA, Huffman KE, **Schwartz JC**, Ram R, Nordsell R, Shames DS, Minna JD, Corey DR, "Involvement of AGO1 and AGO2 in mammalian transcriptional silencing", *Nat Struct Mol Biol.* 2006 Sep;13(9):787-92.
5. Janowski BA, Kaihatsu K, Huffman KE, **Schwartz JC**, Ram R, Hardy D, Mendelson CR, Corey DR, "Inhibiting transcription of chromosomal DNA with antigene peptide nucleic acids", *Nat Chem Biol.* 2005 Sep;1(4):210-5.
6. Janowski BA, Huffman KE, **Schwartz JC**, Ram R, Hardy D, Shames DS, Minna JS, Corey DR, "Inhibiting transcription of chromosomal DNA with antigene RNA", *Nat Chem Biol.* 2005 Sep;1(4):216-22.
7. Lu X, Hu Z, **Schwartz J** "Phase Transition Behavior of Hydroxypropylcellulose under Interpolymer Complexation with Poly(acrylic acid)" *Macromolecules* 2002; 35(24): 9164-8
8. F. D. McDaniel, B. L. Doyle, C.H. Seager, D.S. Walsh, G. Vizkelethy, D.K. Brice, C. Yang, P. Rossi, M. Nigam, M. El Bouanani, G.V. Ravi Prasad, **J. C. Schwartz**, L.T. Mitchell, and J.L. Duggan, "Ionoluminescence decay measured with single ions" *Nuclear Instruments and Methods in Physics Research B,* 2002; 190:1-10.

LIST OF FIGURES

CHAPTER ONE: RNA interference and transcriptional gene silencing

1.1 RNA INTERFERENCE

Double stranded RNAs can silence gene expression through RNA interference (1, 2). The collection of mechanisms by which RNA controls gene expression by targeting mRNA is known as post-transcriptional gene silencing (PTGS). The phenomenon of gene silencing by RNA was initially observed in the early 1990's in *C. elegans* (*3*) and also in the plant *Nicotiana tabacum* (*4*).

In *C. elegans*, the gene *lin-14* was found to be regulated by a gene *lin-4* (*3*). Loss of *lin-14* and *lin-4* has opposite phenotypes. Also, two consecutive stages of *C. elegans* development are characterized by high then almost undetectable levels of LIN-14 protein but no change in the mRNA. A study of the gene products from *lin-4* yielded two short noncoding RNAs, one 61 and one 22 nucleotides long. The longer forms a hairpin structure. The shorter RNA is derived from the longer. The short RNA is complementary to several sites within the *lin-14* 3'-UTR. This model proposes that *lin-4* short RNA binds antisense to the 3'-UTR of *lin-14* mRNA and inhibits protein translation.

Fire et al., in 1998, characterized RNA interference (RNAi) in *C. elegans* (5). Fire et al. set out to characterize the sequence specific knockdown of gene expression by injecting *C. elegans* with single stranded RNAs antisense to their target mRNA. Fire et al. also tested sense RNAs and double stranded RNAs. Double stranded RNAs proved to produce lasting silencing of target gene mRNA and reduced protein levels far more effectively than single stranded RNA. Fire et al. tested the possibility for double stranded RNA to silence gene expression by targeting gene promoters or introns but observed no activity (*5*). Because of the potency of this knockdown, Fire et al. predicted an enzymatic component to this activity. Thus, a model was proposed that double stranded RNA induced the sequence specific silencing of the mRNA.

A subsequent screen of *C. elegans* mutants deficient in RNAi led to the discovery of protein machinery involved, such as argonaute proteins (*6*). Argonaute proteins are a class of RNA-binding proteins that bind a short RNA strand. Argonautes use the short RNA as a template to bind other complementary RNAs. Upon binding

argonaute possess an RNaseIII-like domain that in many cases will cleave the RNA target. The RNA interference phenomenon was also found to be mediated by Argonaute proteins in plants (*7*) and in *D. melanogaster* (*8*).

The early RNAi studies used long double stranded RNAs on the order of hundreds of nucleotides long (*5*). The mechanism by and the form in which these long RNAs would interact with their target genes was obscure. The link between RNA interference and the first sighting of RNA silencing by short 22 nt antisense RNAs in *C. elegans* (*3*) was unfortunately overlooked, likely due to the intense focus on fully complementary duplex RNAs. From plants, it was noticed that PTGS-inducing transgenes gave rise to production of 25 nucleotide long antisense RNAs complementary to mRNA of the silenced gene (*9*). These were speculated as a possible mechanism for specificity for RNAi. Then, in *D. melanogaster* embryo lysates, long double stranded RNAs were found to be digested into 21 and 23 nt duplex RNAs and mRNA targets were similarly degraded into 21 to 23 nt fragments, demonstrating that RNAi is in fact mediated by short 21 nt duplexes directing specific cleavage of the mRNA target (*10*).

In mammalian, double stranded RNA was found to induce RNAi by injecting mouse embryos (*11*). Then 21 nt duplex RNAs transfected using cationic liposomes induced potent RNA interference in mammalian cell culture (*12*). These small RNAs were referred to as small interfering RNAs or siRNAs. As with other organisms, mammalian RNAi also involves argonaute proteins (*13*). Since then, RNAi has become a standard tool for sequence specific knockdown of gene expression in cell culture (*14*).

RNAi has been induced by duplex RNAs that are fed, injected, or transfected into animals, lysates, and cell culture. However, RNAi is not only a tool for researchers. Endogenously expressed small RNAs have been discovered that mediate RNA interference (*15-17*). *lin-4* discussed above is an example of one small RNA capable of mediating RNA interference. These small RNAs are not fully complementary to their targets. They silenced translation without the marked degradation of the mRNA transcript seen with siRNAs. These endogenous RNAs are called microRNAs or miRNAs. Since the discovery of miRNAs, several more subcategories of small RNAs have been characterized that also regulate gene expression (*18*).

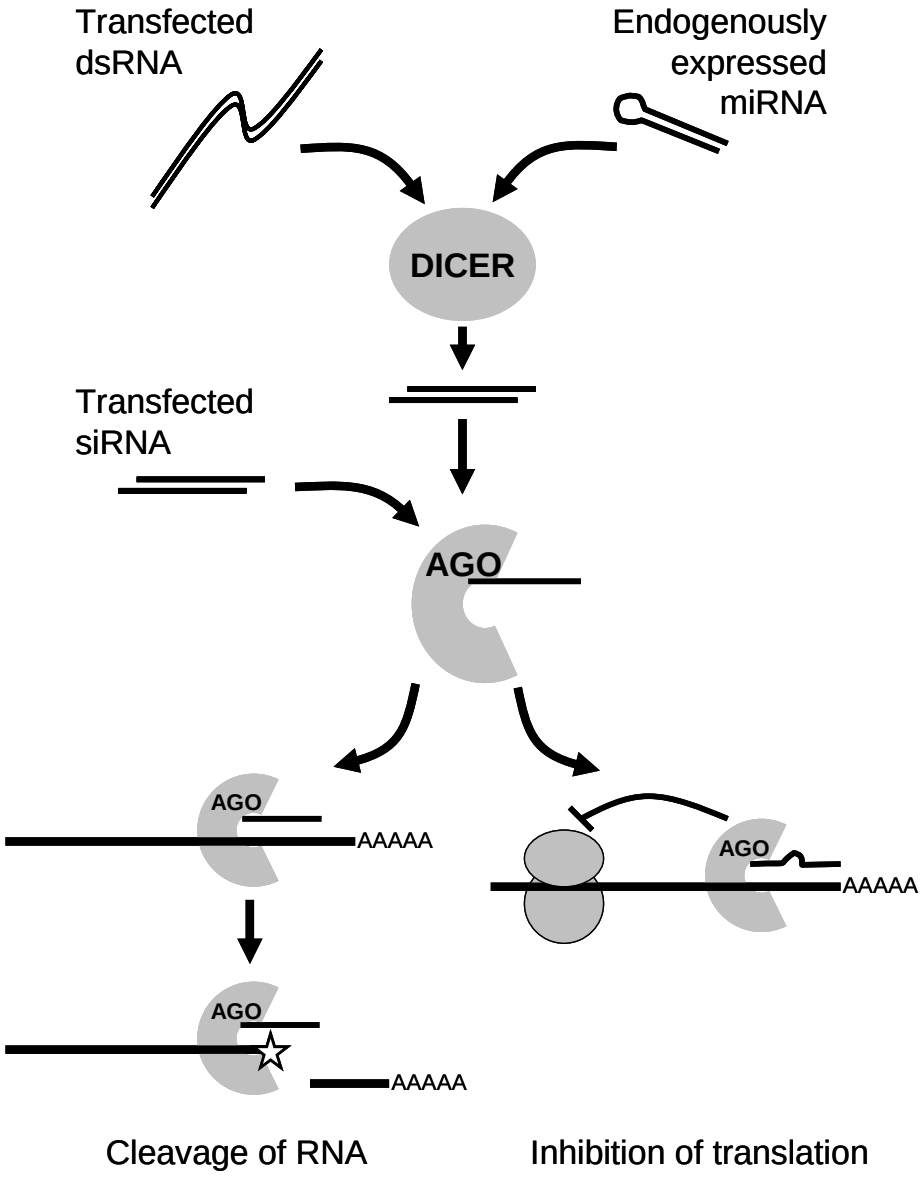


Figure 1: Outline of post-transcriptional gene silencing. Double stranded RNA that is exogenously introduced or endogenously expressed as a hairpin is loaded into argonaute proteins, which binds its complementary RNA target to either cleave the RNA or inhibit translation.

1.2 TRANSCRIPTIONAL GENE SILENCING IN PLANTS

DNA can be transformed into plant leaf disks that are incubated with *Agrobacterium tumefaciens* expressing a tumor inducing plasmid. To insert a gene into plant progeny, a tumor inducing plasmid is constructed containing a gene of interest (*19*). Multiple genes can be co-transformed in this manner, including antibiotic resistance markers. Progeny that sprout from leaf disks are selected by antibiotics. Leaf disks can be prepared from these progeny and a second round of transformation can be performed.

The earliest report of gene silencing in 1989 in plants described that two sequential tranformations of the same transgene induces silencing of both transgenes (*19, 20*). This phenomenon was originally called "homology-dependent gene silencing" and later became known as transcriptional gene silencing (TGS). TGS, in plants, occurs when gene promoter DNA cytosines are 5-methylated. Methylation is induced when there exists homology between the promoters of the host gene and the transgene (*4, 19-22*). Interestingly, PTGS mechanisms in plants also induces DNA methylation at the 3' ends of genes in plants (*21*). Models to explain this mechanism included DNA-DNA interactions whereby the two DNA copies found each other and were silenced.

The idea that DNA methylation is induced by RNA came from the observation that during viroid replication, the production of the highly structured RNA genome correlated with methylation of cDNA copies of the viroid genome (*23*). Motivated by this, expression of an RNA complementary to a gene promoter sequence was shown to induce methylation and silencing of the targeted promoter. This demonstrated that TGS depends on RNA transcripts being expressed from the promoter of transgenes (*21*). An observation was made that DNA methylation was strictly confined to areas complementary to the RNA transcript, which was taken to suggest mediation by an RNA-DNA hybridization event (*21, 23*). DNA methylation was found to be dependent on the formation of double stranded RNA, which implicated involvement of the RNAi pathway (*24*). Even though RNAi is usually studied in the context of RNA-RNA interactions, even recent literature considers the possibility that RNA-directed DNA methylation (RdDM) could be mediated by RNA-DNA interactions (*25, 26*).

The RNAi machinery in plants makes a distinction between short RNAs that mediate TGS and those that target PTGS. Two classes of small RNAs are produced from double stranded RNAs in plants: short 21 to 22 nt long RNAs and longer 24-26 nt long RNAs. By transfecting different viral suppressors of RNAi locally into plant leaves, some can silence the production of longer small RNAs and some suppress the production of all small RNAs. Using this technique, it could be extrapolated that short 20-21 nt RNAs mediate mRNA degradation and PTGS but not DNA methylation. Longer 24-26 nt RNAs mediate DNA methylation and TGS but not mRNA degradation (*27*).

Plants have 10 argonaute proteins, of which ARGONAUTE1 is involved in siRNA mediated degradation of mRNA and miRNA activity (*28*). The earliest studies found *A. thaliana* ARGONAUTE4 (*a*AGO4) responsible for RNA-directed DNA methylation (*29*). However, at some other loci, other argonautes may be involved (*25*). Interestingly, *a*AGO4 is a catalytically active argonaute protein capable of using a small guide RNA to direct the protein to cleave a target RNA molecule (*30*). The ability for *a*AGO4 to methylate DNA does not seem dependent on its catalytic activity. This is evident when *a*AGO4 is mutated to be inactive. Exogenously produced small RNAs can still induce DNA methylation and TGS. The production of endogenous small RNAs that induce DNA methylation does seem dependent on the catalytic activity of *a*AGO4 (*30*).

A question about how TGS in plants is maintained concerns endogenously produced small RNAs. These work in *cis* to induce methylation of their own coding loci and silence RNA polymerase II (Pol II) leaving no mechanism for continual small RNA production for reinforcement of DNA methylation and silencing. This problem was resolved by the discovery of RNA polymerases IV (Pol IV) (*31, 32*) and V (Pol V) (*33*) (originally referred to as RNA Polymerase IVb) that work in concert to reinforce DNA methylation. These polymerases are not regulated the same manner as RNA polymerase II and thus can work to produce substrates for small RNA production and nascent transcripts for argonaute recruitment in otherwise transcriptionally silent loci. Further analysis revealed that Pol IV and Pol V were essentially modified Pol II enzymes, specialized for their roles in production of small RNAs and noncoding RNAs (*34*).

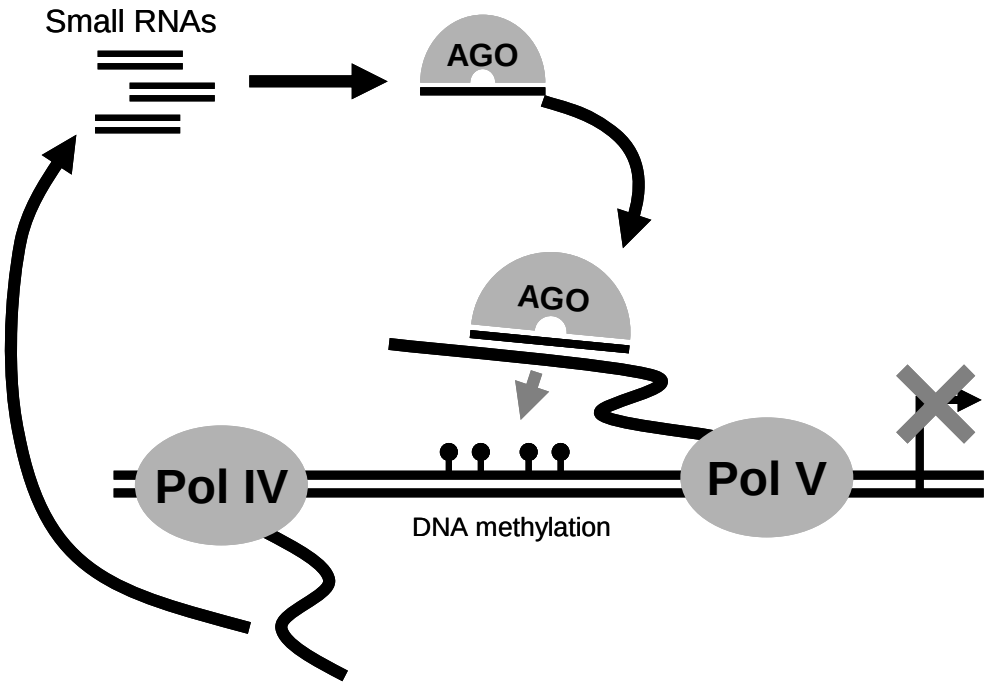


Figure 2: Transcriptional gene silencing in plants. RNA polymerase IV is required for production of double stranded RNAs that then recruit Ago4 to RNA polymerase V transcripts, inducing methylation of genomic DNA and silencing RNA polymerase II transcription.

Current evidence suggests that Pol IV is responsible for production of small RNAs that are bound by *a*AGO4 (*32*) and then *a*AGO4-RNA complexes are recruited to DNA by Pol V, which produces nascent RNA transcripts to serve as a substrate for *a*AGO4 to bind (**Fig. 2**) (*35, 36*). Some other co-factors that interact with Pol V must somehow determine that in some cases robust transcriptional silencing is observed along with DNA methylation and, in other cases, less robust silencing and reduced DNA methylation is observed but still with the silent chromatin markers such as methylation of lysine 9 of histone 3 (H3K9) (*37*).

1.3 TRANSCRIPTIONAL GENE SILENCING IN *S. POMBE*

Concurrent with research characterizing TGS in plants was the discovery of TGS in the single celled eukaryote, fission yeast or *S. pombe*. The two most studied yeast organisms, *S. cerevisiae* and *S. pombe*, have diverged in their conservation of the RNAi machinery. *S. cerevisiae* has lost its argonaute protein and its RNA-dependent RNA polymerase. Unlike other eukaryotes that have multiple argonaute proteins, *S. pombe* only has one argonaute (*sp*Ago1) protein and a RNA-dependent RNA polymerase (*sp*Rdp1) (*38*). *S. pombe* has retained their Dicer proteins (*sp*Dcr1) (*39*), which are required for the specific cleavage of 21 or 23 nt long RNAs from longer double stranded RNAs. In fact, most budding yeast studied, except for *S. cerevisiae*, have retained a *Dicer* and *Ago* gene and it is important for the silencing of retrotransposons. It has been speculated that mutation of retrotransposons to inactivity in *S. cerevisiae* may have negated the need for RNAi machinery in this species (*40*).

In contrast to plants, yeast does not methylate its DNA (*25*). Instead, yeast transcriptional silencing is associated with methylation of histones, particularly the 9th lysine on the N-terminal tail of the histone 3 (H3K9), to induce heterochromatin. Inspired by work in plants and *D. melanogaster*, knockouts in *S. pombe* were generated for the three RNAi proteins, *sp*Ago1, *sp*Dcr1, and *sp*Rdp1 to test if heterochromatic silencing was effected. The result was loss of transcriptional silencing, acculumation of transcripts originating from heterochromatic regions, and loss of H3K9 methylation (*39*).

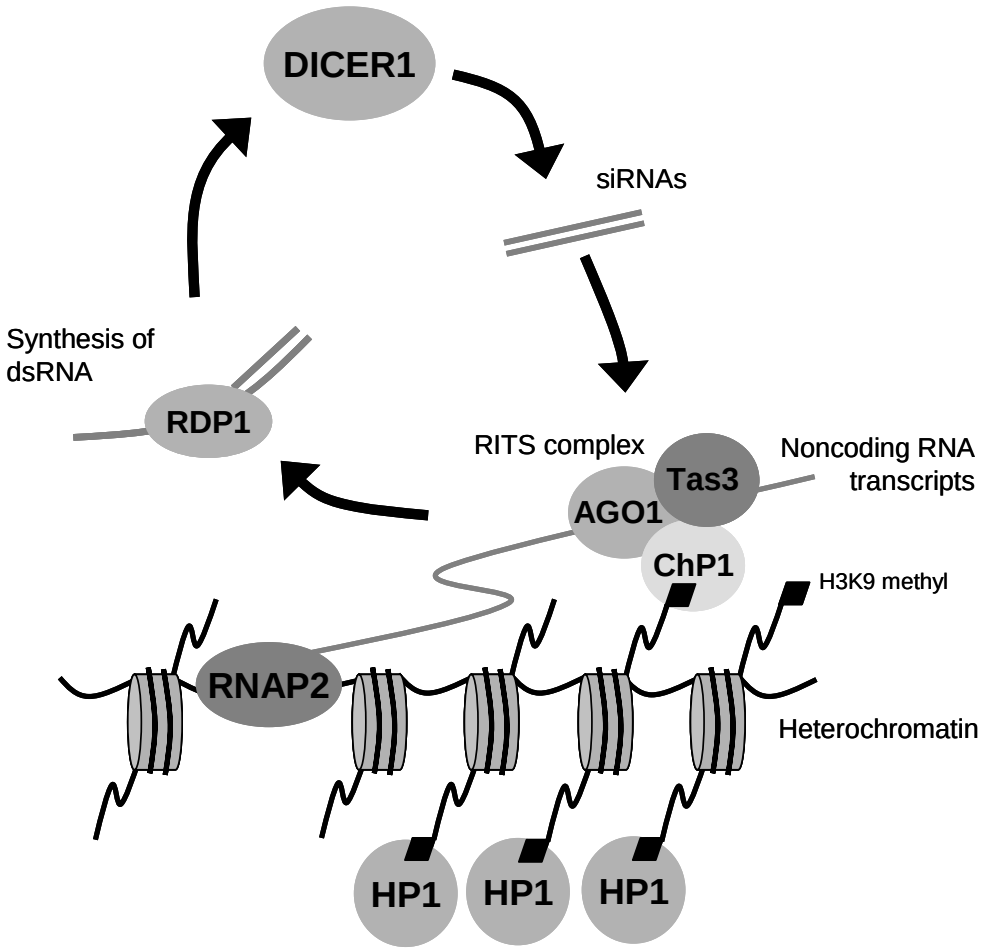


Figure 3: Transcriptional gene silencing in *S. pombe*. RNA transcription through heterochromatic regions leads to the production of double stranded RNA that is cleaved by Dicer into siRNAs. These siRNAs are loaded into the RITS complex and recruits RITS to heterochromatin. Recruitment of RITS

A protein complex was purified with *sp*Ago1 and found essential for TGS in *S. pombe*. This complex included the argonaute protein *sp*Ago1, a chromatin binding protein Chp1, and a novel protein with unknown function Tas3 *(41)*. Chp1 contains a chromodomain that is known to bind methylated H3K9 tails. The proteins of this complex were found to be highly enriched in heterochromatic regions of the genome (*42*). This complex interacted with the *sp*Rdp1 complex and with noncoding RNAs expressed from heterochromatin regions of the yeast genome (**Fig 3**) (*43*).

The discovery of the RITS complex had two novel implications. Not only is RITS a novel RNAi-associated TGS complex but its recruitment to heterochromatic regions is dependent on transcription of heterochromatic regions. This was a difficult result to reconcile since heterochromatin is traditionally thought of as being transcriptionally silent. However, it was shown that by creation of a fusion protein with λN-peptide to Tas3, and insertion of λN's cognate RNA binding sequence in an *ura4*$^{+}$ gene, heterochromatic TGS could be induced (*44*). This construct isolated the role of the RNA transcript serve as a target for heterochromatic TGS, rather than its parallel role as the template for double stranded siRNA production. Active transcription of heterochromatic silenced DNA was confirmed by chromatin immunoprecipitation of RNA polymerase II and by nuclear run on assay (*44, 45*).

As mentioned above, RITS activity in *S. pombe* is closely associated with chromatin modifications. Unfortunately, the transcriptional signals associated with these chromatin modifications have appeared to diverge throughout eukaryotes making it difficult to extrapolate those observations to human. RITS induces di- and trimethylation of H3K9 (H3K9me2 and H3K9me3) in *S. pombe* to signal for transcriptional silencing. However, in animals including *D. Melanogaster* and human, heterochromatic DNA associated with H3K9me3 can result in either gene silencing or gene activation through mechanisms that remain to be understood (*46*). Similarly, other chromatin marks have confusing outputs even in *S. pombe*. Histone deacetylation can reverse RITS induced heterochromatin formation and induce gene activation. This is the result of silencing expression of the repeat associated noncoding RNAs that serve as the source and targets for RITS associated siRNAs. On the other hand, in many euchromatic regions acetylation

is required for gene expression, thus histone deacetylase complexes silence gene expression (*47*).

RITS in heterochromatin targets in *cis*, producing siRNAs that target the region they are expressed from. RITS activity was shown to be able silence a euchromatic region in *trans*. RNA hairpins were expressed from a construct incorporated in one region of the genome. These RNAs were complementary to an *ura4*$^{+}$ construct incorporated into either a heterochromatic or euchromatic region of the genome. These RNAs could induce transcriptional silencing in the heterochromatic locus. They could also induce silencing in the euchromatic locus if the heterochromatin associated protein HP1 was overexpressed (*48*). Without overexpressed HP1, the small RNAs merely induced PTGS. This suggested that TGS was possible in euchromatic regions but was dose dependent on having enough of the necessary protein machinery nearby and enough small RNAs.

1.4 piRNA ASSOCIATED TRANSCRIPTIONAL GENE SILENCING IN *D. MELANEGASTER* AND MAMMALS

Flies have not been directly shown to have TGS mechanisms of the same type that have been shown in plants or yeast (*49*). However, some of the earliest evidence of RNA induced TGS was shown in flies (*50, 51*). TGS in flies occurs in germline cells that primarily silences transposable elements during earliest stages prior to embryogenesis. If TGS occurs in somatic cells, convincing evidence of this remains to be demonstrated.

In germline, a novel group of argonaute proteins named after its founding member, Piwi, has been shown to be required for stem cell self-renewal and gametogenesis (*52-54*). Mutations in the related Piwi protein, Aubergine, was shown to cause sterility in males (*51*). Additionally, double stranded RNA was shown to be required for silencing of the *D. melanegaster* repeat elements *Stellate* and *Supressor of Stellate*. Silencing of these elements requires interactions with Aubergine *(50, 51)*. Mutations in Piwi derepresses the retrotransposon *gypsy (55)*. Mutations in Aubergine derepresses the telomeric retroelement *TART (56)*. Finally, *aubergine* mutants also have derepressed P-element transposition (*57*).

Taken together these data raised the expectation that piwi proteins would be involved in some RNAi related silencing of repeat elements and possibly heterochromatin maintenance (*54*). Also, libraries of small RNAs sequenced from *D. melanogaster* revealed a novel group of RNAs derived from repeat sequences (*58*). Whereas miRNA and siRNA sizes clustered around 19 to 22 nucleotides long, these repeat associated RNAs cluster around sizes of 24 to 27 nucleotides long (*58, 59*). Knockout of piwi proteins Piwi or Aubergine caused derepression of the mRNA targets of the longer RNAs but knockout of standard RNAi proteins Dicer1, Dicer2, Ago1, or Ago2 had no effect (*59*). Two mysteries remained, what enzyme was responsible for synthesis of these novel RNAs and why the high propensity for the 5' nucleotide of these repeat associated RNAs to be a U (*54, 60*). The latter mystery remains unsolved.

Parallel to the characterization of piwi proteins in *D. melanogaster* was the characterization of their mammalian homologues in mouse germline cells (*52, 61*). The proteins, Mili and Miwi, were immunoprecipitated and their associated RNAs sequenced to reveal a novel family of mammalian RNAs from 25 to 30 nucleotides long (*62, 63*). The same results were seen in rat (*64*). These RNAs clearly differed from standard siRNAs or miRNAs which tend to be 21 to 22 nucleotides long. Unlike the repeat associated RNAs from *D. melanogaster*, these RNAs were not particularly enriched for repeat element sequences but tended to be derived from distinct clusters one the mammalian genome. These piwi associated RNAs were christened piRNAs.

Revisiting the *D. melanogaster* system, antibodies to each of the piwi proteins, Piwi, Aubergine, and Ago3 were developed. These antibodies were used for immunoprecipitation and the associated RNAs were sequenced. The results were the repeat associated RNAs. Thus, repeat associated RNAs are piRNAs (*65*). However, aligning sequences with the genome revealed a bias in that Piwi and Aubergine associated RNAs are expressed from one strand of the DNA, Ago3 RNAs mapping to the same location will be exclusively expressed from the opposite strand. Also, whereas Piwi and Aubergine associated piRNAs have more than 70% of their 5' nucleotides as a U, Ago3 is strongly biased to associate with RNAs that have an A at position 10 (*65, 66*).

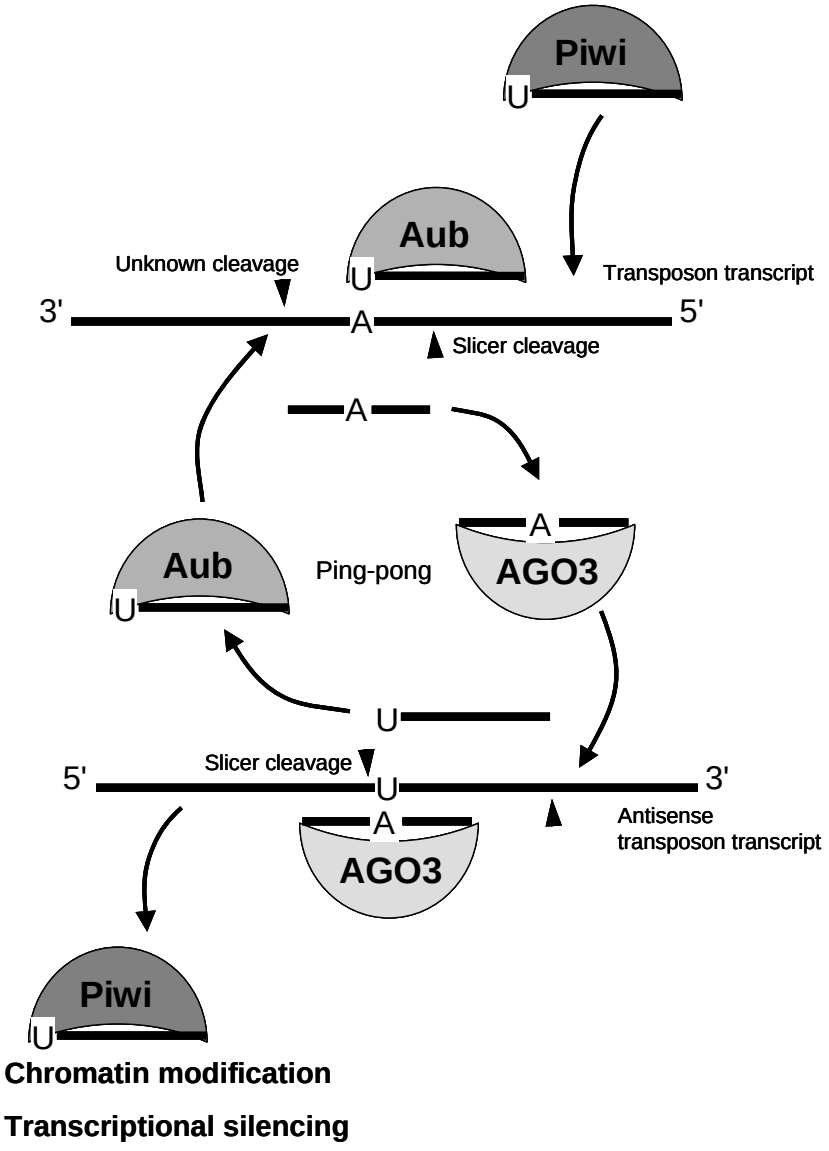


Figure 4: piRNA transcriptional gene silencing. The ping-pong cycle is hypothesized to use sense and antisense transcripts for generation of short RNAs in a Dicer independent fashion that are loaded into Piwi. These short RNAs are used to silencing transcription and direct heterochromatin formation.

The bias at position 10 suggested that perhaps Ago3 associated RNAs were the cleavage products from Piwi and Aubergine associated RNAs. This model is known as the ping pong cycle (**Fig 4**). An antisense RNA is expressed from a region and loaded into a piwi protein. This catalytic protein finds a sense transcript, binds, and cleaves it. The cleavage product is loaded into Ago3 and the process continues. Presumably differences in the relative catalytic activity of Piwi, Aubergine, and Ago3 account for the relative abundance of 5' U RNAs in comparison to RNAs with an A at position 10 (*54*). In both mouse and *D. melanogaster*, piRNAs are associated with chromatin modifications, DNA methylation, and heterochromatin silencing of repeat elements (*49, 67, 68*).

1.5 agRNA INDUCED TRANSCRIPTIONAL SILENCING IN HUMAN

In 2004, literature about plant TGS and fly TGS was still very confusing and would remain so for another couple years. *S. pombe* TGS was just being characterized and the RITS complex had been purified and its components identified (*41*). During this time a brief paper appeared suggesting that a 21 nucleotide long, double stranded RNA (identically designed as standard siRNAs) targeting a gene promoter could induce transcriptional gene silencing in human cells (*69*). In this case the target gene was a lentivirally inserted GFP behind an elongation factor 1A (EF1A) promoter in Hela cells.

For this experiment one duplex targeting the EF1A promoter, one duplex targeted the GFP mRNA, and one control duplex was tested. mRNA levels were measured by quantitative real time PCR revealing reduction of GFP expression. Nuclear run on assay, which measures production of nascent RNA transcripts from active, chromatin bound polymerases isolated from live cells, showed reduction of GFP transcription. Cleavage of genomic DNA with a methylation specific restriction enzyme suggested that the EF1A promoter was methylated. The EF1A promoter duplex also was shown to silence endogenous EF1A expression and restriction enzyme digestion suggested promoter methylation. The silencing was reversed by treatment with the histone deactylase

inhibitor, tricostatin A (TSA) in combination with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytosine (5-aza-C).

Unfortunately, this paper left much to be desired in the way of widely accepted controls. The pitfalls of oligonucleotide transfection associated off-target effects had already been well documented before this paper (*70-73*). Problems include that 1) only one active sequence was shown, 2) the active sequence was in no way mutated with mismatches or scrambles to test for sequence specificity, 3) a dose dependent response or time course was not shown, 4) effects were only shown at a high concentration of 100 nM raising the possibility of off-target effects, and 5) since the endogenous EF1A was silenced as well, the control was missing showing the effects of EF1A knockdown by standard siRNA targeting the mRNA. This last positive control is important because for the gene RASSF1A, DNA methylation was seen for both promoter targeting RNAs and siRNAs suggesting the effect is a downstream effect of RASSF1A knockdown .

One month after this paper, a second paper appeared claiming RNA directed DNA methylation in human cells (*74*). Although at first this paper seemed more suffisticated and more controlled, it ultimately proved to appear fraudulent and was retracted. Ten duplex RNAs were tested targeting the promoter of the gene e-cadherin in human breast cancer MCF7 cells. No duplex transfected at the high concentration of 100 nM could reduce e-cadherin mRNA or protein expression more than 50 %. Instead, all ten duplexes were transfected at once and potent knockdown was seen.

In 2005, another paper appeared concerning RNA directed TGS in human cells (*75*). This time the target gene was RASSF1A. Duplex RNAs were produced by retrovirally inserting a construct expressing an RNA hairpin into the genome. These RNAs are processed by the miRNA pathway including the proteins Drosha and Dicer to produce double stranded RNAs that are then loaded into the RNAi pathway (*76, 77*). Moderate knockdown of the mRNA was shown by semi-quantitative PCR where the amount of RNA is gauged by the density of bands from PCR product seen by agarose gel electrophoresis and staining by ethidium bromide. DNA methylation was measured by modern bisulphate treatment. Three sequences mutated to include four mismatches each were employed as negative controls.

The promoter targeting duplex RNA slightly reduced RASSF1A mRNA expression. Bisulfate sequencing revealed methylation of DNA in the RASSF1A promoter. DNA was not methylated in the cells treated with mutated RNA sequences showing sequence specificity. However, DNA was also methylated when RASSF1A was silenced by standard siRNA targeting the mRNA. This result can be interpreted that the promoter methylation is due to loss of RASSF1A expression and not due to RNA directed DNA methylation at the promoter.

Finally, in 2005, a paper appeared studying double stranded RNAs targeting the promoter of the viral gene HIV-1 in HIV infected human MAGIC-5 cells (*78*). In this case long term silencing of gene expression was observed extending to 38 days for two RNA duplexes targeting the HIV-1 promoter. Both methylation specific restriction enzyme treatment and bisulfite sequencing revealed DNA methylation. Of three negative controls – mock, mRNA targeting siRNA, or scrambled sequence – only the scrambled sequence induced small amounts of methylation of the HIV-1 promoter seen by bisulfite sequencing but not by methylation specific restriction enzyme treatment. 5-aza-C treatment only partially reversed silencing of the HIV-1 promoter suggesting that DNA methylation may not be required for transcriptional silencing.

This study also could have done more to establish the possibility of mammalian TGS. Knockdown of HIV-1 is only shown by western and quantitative RT-PCR for one of the duplexes targeting the promoter. The authors did perform a nuclear run-on assay to demonstrate silencing at the level of transcription. However, only one subsection of one figure in the paper has error bars and those look to be technical replicates rather than repeated experiments. The nonspecific methylation of the HIV-1 promoter by the scrambled RNA sequence is concerning but the effects was far less than the amount of methylation observed by complementary RNA sequences. More concerning is that one duplex did not knockdown HIV-1 expression but methylation assays say that the DNA is methylated. Another duplex did not knockdown HIV-1 and methylation specific restriction enzyme treatment did not detect methylation but bisulfite sequencing did detect methylation.

Everything considered, the most optimistic outlook from these papers would be to suggest that viral genes may be subject to silencing by methylation of the DNA. Since no other evidence of RNA directed DNA methylation has been observed for endogenous genes, this may be a feature unique to viral genes and not associated with human genes. A more pragmatic outlook would be to assume that evidence of RNA directed DNA methylation has not been shown in mammals. To date, no further evidence supporting the notion of RNA directed DNA methylation of endogenous mammalian genes has been shown and the before mentioned results have never been repeated, even by the same labs that published them.

A final note on RNA directed DNA methylation, a study based on methylation specific restriction digest in Dicer knockout mouse stem cells concluded that RNAi had no role in maintanence of centromeric DNA methylation (*79*). Later two papers appearing in the same journal issue, one from the same lab that originally reported no role for RNAi in centromaric methylation, did find RNAi to be involved in DNA methylation of centromeric DNA in mouse embryonic stem cells (*80, 81*). This control was mediated by a miRNA cluster, miR-290, which controls expression of the gene retinoblastoma-like 2 protein (Rbl2) and Rbl2, in turn, regulates expression of the DNA methyltransferases. Knockout of the gene Dicer in mouse embryonic stem cells silenced production of miR-290 and induced global reduction of DNA methylation. This study highlights the difficulty of distinguishing causes of experimental results when modulating a master regulatory pathway such as the RNAi pathway, which regulates so many genes through PTGS pathways.

1.6 RNA INDUCED ACTIVATION

Recently, papers by Vasudevan et al. have appeared suggesting that an RNAi mechanism may be involved in activation of translation under certain cell conditions. While studying the regulation of the mRNA for human TNFα by a minimal AU-rich element (ARE), it was noted that certain conditions induced increases in translation for a firefly luciferase gene fused to a 3' UTR containing an ARE (*82*). These conditions were

that cells were allowed to grow to saturation under serum starved conditions to induce cell cycle arrest. Alternatively, the activation could be induced by stopping cell cycle with the drug, aphidicolin.

Purification of a biotinylated ARE under serum and serum starved conditions revealed that FXR1 associates with the ARE only under serum starved condition. FXR1 could be immunoprecipitated with the TNFα mRNA under serum starved conditions. shRNA knockdown of several of the splice isoforms of FXR1 reverse activation. FXR1 fused to a λN peptide sequence forces recruitment to the firefly luciferase mRNA fused with a Box B RNA sequence in its 3' UTR and induces activation of luciferase expression even under normal serum growth conditions.

FXR1 is an RNA binding protein known to interact with the argonaute proteins and human Ago2. This interaction was confirmed under serum starved conditions for formaldehyde crosslinked samples only. Also, Ago2 could be immunoprecipitated with the firefly mRNA. Ago2 recruitment to Box B containing firefly luciferase 3' UTRs could be forced by using a λN fused Ago2. Under this condition, translation could be activated only for serum starved conditions and this activation was reversed by shRNA knockdown of FXR1. Similarly, λN fused FXR1 activated Box B fused firefly luciferase mRNA in both serum and serum starved growth conditions but was reversed by shRNA knockdown of Ago2. Furthermore, Ago2 migrated with polysome bound mRNA during serum starvation conditions and otherwise migrated with non-polysome bound mRNA suggesting a general role in translational activation under these conditions.

This paper was followed by a study to identify the miRNA responsible for the ARE dependent activation (*83*). By mutating a miR-369 seed sequences in a firefly luciferase fused to a TNFα 3' UTR, this study suggests that miR-369 is responsible. This paper further suggested that Ago2 dependent activation is more general under the conditions of growth arrest by serum starvation. The 3'-UTR of HMGA2, a target for the miRNA Let-7, fused to firefly luciferase saw activation under serum starved conditions and this was increased further by transfection with exogenous Let-7 RNA. In normal serum conditions exongenous Let-7 RNA transfection induced the expected gene silencing.

A final study from Vasudevan et al. performed cell cycle synchronization and found that activation only occurs during the G1 phase of the cell cycle (*84*). Furthermore, mutation of the 3'-UTR to provide a fully complementary target for the miRNA caused silencing instead of activation. This is taken to suggest that full complementarity induces siRNA cleavage rather than standard miRNA regulation. The authors also suggest that since all genes are not upregulated by this growth arrest conditions that this effect is likely the sum of multiple inputs, of which recognition by miRNPs may not dominate.

A role for miRNAs and RNAi in gene activation is unexpected and even controversial. These papers present quality data and have the good use of controls, even in comparison to other papers in high profile journals, however there are deficiencies that would warrant further investigation. First, all studies were done with transfected plasmids expressing firefly fused 3'-UTRs, meaning that all cells must undergo the stress of transfection. For such a novel claim, it would be desirable to see the effect on an endogenous gene. Second, the effect is only seen under very specific and stressful conditions of confluency induced growth arrest during serum starvation. This makes this system very difficult to reproduce and raises questions about the biological relevance of such a severe treatment of cells. Taken together, the results cannot be disregarded but more investigation is needed to establish the biological significance of this observation.

CHAPTER TWO: Introduction to Noncoding RNAs

2.1 THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

The central dogma of molecular biology, as enumerated by Francis Crick, concerns the "residue-by-residue transfer of sequential information" between the biopolymers DNA, RNA, and proteins (**Fig 5**) (*85, 86*). Simply based on structural considerations for nucleic acids and amino acids, it hypothesizes 1) that DNA can be duplicated into DNA and RNA can be duplicated into RNA, 2) DNA can be transcribed into RNA and RNA could be reverse transcribed into DNA, and 3) RNA through the aide of tRNAs can be translated into protein. At the time that the dogma was put forth, no evidence of reverse transcription or RNA dependent RNA polymerization existed, but the process could not be excluded due to the structural similarity of the DNA and RNA polymers. However, transfer of protein to either DNA or RNA was deemed as "highly unlikely" because of the structural complexity of the protein polymer (*85*).

Although the central dogma makes no claims about the molecular mechanisms for transfer from one biopolymer to another, the central dogma is laid out like a spectrum for molecular biology ranging from information preservation, handled by the DNA molecule, to biochemical catalysis, handled by proteins. Early in molecular biology, RNA was regarded as the passive intermediate that faithfully conveyed the code of the DNA to produce the corresponding amino acid sequence for the biological business of enzyme activity and molecular signaling. The special exception was the RNA virus where RNA served as the mediator of genetic preservation (*87, 88*). Then at other end of the biological spectrum, it was discovered that RNA can catalyze chemical reactions. This was initially observed in two instances: RNA could without a protein co-factor catalyze splicing of an intron (*89*) and RNase P could catalyze cleavage of tRNA precursor during their maturation (*90*).

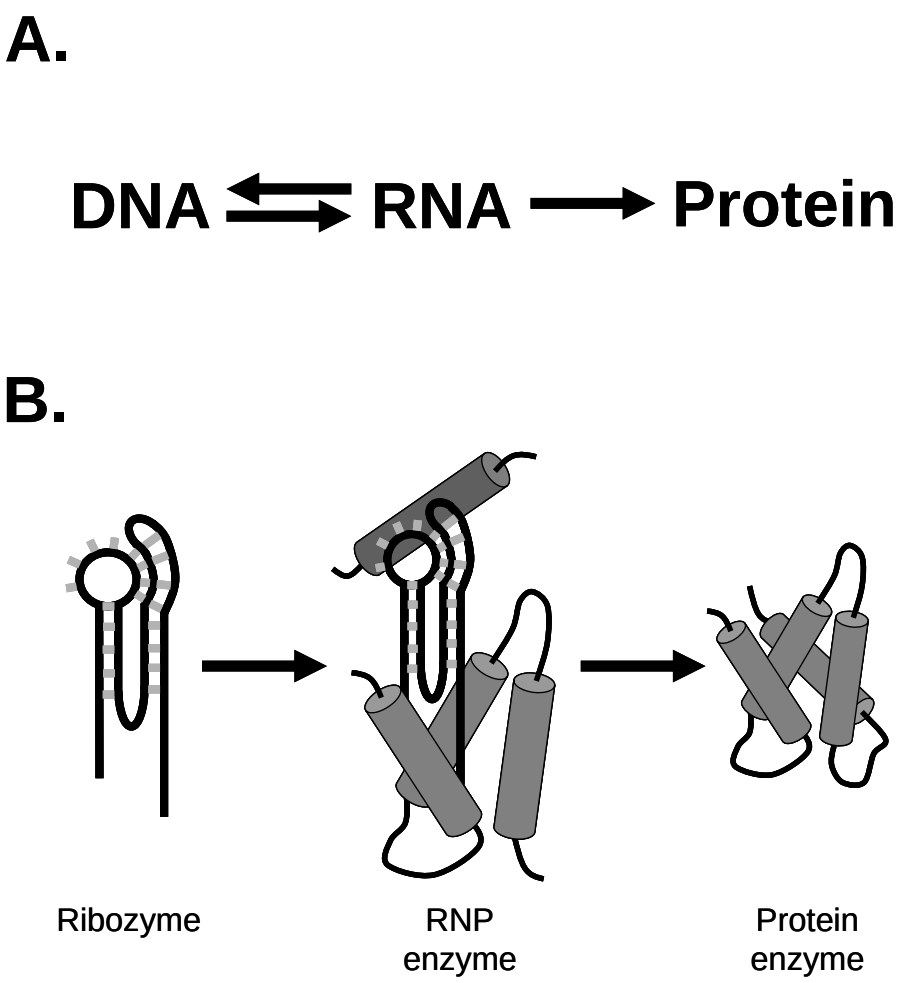


Figure 5: The central role of RNA. (A) The central dogma of molecular biology states that DNA is transcribed into RNA and RNA is translated into protein. Also, RNA can be reverse transcribed into DNA. (B) The RNA world suggests that in early molecular evolution RNA was the catalytic engine and genetic material of life. Protein co-factors and protein enzymes evolved later to take a more central role in catalysis. Today, all three forms of catalytic molecules exist: ribozymes, RNP enzymes, and protein enzymes.

2.2 THE RNA WORLD

The ability of RNA to both serve as a mode of genetic inheritance and as a catalytic engine suggests a possibility that some early point in molecular evolution, the spectrum of life's molecular biology may have been entirely populated by RNA molecules. As with the central dogma of molecular biology, this hypothesis of an "RNA world" was made by Francis Crick by considering the role of RNA in the ribosome for protein synthesis (*91*). Evolutionary biologists have and continue to speculate that modern enzymes that require single nucleotides or small RNAs as a co-factor might point back to an era when the role for the protein enzymes in life's chemistry may have been diminished to the point of non-existence (*92*).

Taken together, the diversity of roles for RNA molecules suggests that even though RNA is often thought of as a medium for executing orders for protein synthesis encoded in DNA, any newly discovered RNA molecule could be involved in any number of processes ranging from co-factors, to signaling molecules, to actual ribozymes (**Fig 5**) (*93*). In fact, fewer than 3% of all RNA in cells are protein encoding messenger RNAs (*94*). These RNAs not involved in transmitting messages from the DNA are termed noncoding RNAs.

2.3 DIVERSITY OF ROLES FOR NONCODING RNAS

2.3.1 Ribosomal RNA

The earliest established role for ribonucleic acids was actually as a noncoding RNA. It was observed that the vast majority of RNA in cells (>90 %) was found in the form of ribonucleoprotein particles known as the ribosome (*95*). These particles comprised largely (~ 66%) of ribosomal RNA (rRNA) were shown to be the sites of protein synthesis (*96*). The eukaryotic ribosome has two subunits with 3 noncoding RNA molecules in the large subunit (28S, 5.8S, and 5S) accompanied by 50 proteins and 1

noncoding RNA molecule in the small subunit (18S) accompanied by 33 proteins. The ribosome is a peptidal-transferase ribozyme that consumes aminoacyl-tRNAs to produce proteins using a messenger RNA template (*93, 97*).

2.3.2 Transfer RNA

In addition to the ribosome being involved in protein synthesis, transfer RNAs (tRNAs) deliver chemically bound amino acids to the to the ribosome for continued peptide chain production (*98, 99*). Whereas around 80% of total mammalian RNA is rRNA, another 15% is tRNA (*100*). Aminoacyl-tRNAs are loaded into the ribosome in an order prescribed by their 3 nucleotide anticodons, which recognize their complementary sequence in the messenger RNA (*91*). The amino acid is then transferred to the growing peptide chain and the free tRNA is released. More than one codon results in the incorporation of the same aminoacid into a protein due to the fact that the third nucleotide does not have the opportunity to perfectly base pair and thus wobbles. Thus, for example in human some 497 tRNA genes produce tRNAs with 48 different anticodons that are loaded with their corresponding 20 aminoacids (*94*).

2.3.3 Small nuclear RNA

Another class of noncoding RNAs is small nuclear RNA (snRNAs) that are required for splicing (*101*). Ultimately, five snRNA molecules were found in eukaryotes that aide in the spliceosome to catalyze the correct splicing of mature messenger RNA from pre-mRNA (U1, U2, U4, U5, and U6). These RNAs, along with some 50 proteins, aide in the ATP-dependent rearrangement of RNA-RNA, RNA-protein, and protein-protein interactions to direct splicing (*94*). ATP is not required for splicing itself but is necessary to drive the restructuring of the spliceosome. More recently, a second spliceosome has been found in metazoans with its own host of snRNAs (U11, U12, U4atac, and U6atac) (*102*). This second spliceosome is present at a much lower abundance and directs certain rarer and non-canonical splicing events.

2.3.4 Small Nucleolar RNA

After discovery of the U-rich small nuclear RNAs involved in splicing, it was found that certain of these RNAs specifically associated with rRNA and not with messenger RNA (*103, 104*). In addition to being specialized in the splicing of rRNAs, they are organized into a subnuclear compartment called the nucleolus. For this reason, these noncoding RNAs were rechristened as small nucleolar RNAs or snoRNAs (*103*). snoRNAs also direct the correct folding of rRNAs and modifications of the rRNAs and of some other RNAs such as the U6 snRNA (*94*). Another unique feature of snoRNAs is that many are not transcribed from their own promoters but are expressed from introns of mRNAs and spliced out (*105*). Like rRNAs that are expressed in every living organism, the snoRNAs are thought to be truly ancient noncoding RNAs with their functional equivalents appearing in both eukaryotes and archeae (*106*).

2.3.5 RNase P

The generation of mature tRNA from pre-tRNA by the removal of a 5' leader sequence is accomplished by the ribozyme RNase P (*107*). In eukaryotes this ribozyme complex is made of one catalytic RNA, H1, and 10 protein cofactors (*108*). In bacteria RNase P requires only one protein cofactor (*109*). At high salt concentrations, the H1 RNA is able to cleave its substrate without the aide of any protein cofactors.

2.3.6 MicroRNAs

As described in chapter 1, the smallest of noncoding RNAs are microRNAs (*110, 111*). These RNAs can be expressed within the introns of coding mRNAs or from their own noncoding RNA precursors. Imperfectly matched hairpin structures are cleaved into miRNA precursors that range from 30 to 60 nucleotides long by the enzyme Drosha. This pre-miRNA is further cleaved to a 20 nucleotide duplex by the enzyme Dicer. Following that the duplex is loaded into the RISC complex, the second strand is removed, and RISC

uses this short RNA as a guide to sequence specifically recognize mRNA targets and repress translation (*112*).

2.4.5 Telomerase RNA

The ends of chromosomes are maintained by a ribonucleoprotein complex called telomerase. The RNA component of telomerase is required for telomere repeat synthesis and contains an RNA sequence, which upon its discovery was immediately thought of as a potential template for the synthesis of telomeric repeats (*113, 114*). Comparison of sequences of various species' RNA sequences or mutated RNA sequences and corresponding telemoric sequences led telomerase to be classified as a reverse transcriptase (*115*), but this was not demonstrated until much later (*116*). Telomerase RNA is more than a template but also flexible scaffold bringing together the protein members of the telomerase complex (*117*). Recently, it has been shown that a conserved pseudoknot structure allows a protruding 2'-OH group to contribute to catalysis raising the possibility that telomerase may actually be a ribozyme (*118*).

2.4 NOVEL ROLES FOR NONCODING RNAS

In addition to the major classes of noncoding RNAs described above, many noncoding RNAs have been described that do not easily fall into simple categories. A common theme in these noncoding RNAs, from bacteria to human, is a role in the regulation of gene expression. Below are a few prominent examples of these novel roles but is certainly not an exhaustive list of all noncoding RNAs under investigation to date.

2.4.1 sRNAs

Recently, computational studies using the *E. coli* and *S. enterica* genomes predicted hundreds of small noncoding RNAs in the intergenic regions of these prokaryotes (*119-121*). Northern analysis and 5'RACE studies confirm that a majority of

these noncoding RNAs are expressed. These small RNAs, sRNAs, range from 50 to 250 nucleotides. Despite the fact that sRNAs share a common name, there is little if any similarity in their mechanism of action. sRNAs bind their complementary RNAs or protein cofactors to either silence or activate transcription or translation (*120, 122*). Most of the time, sRNAs silence expression. Many RNAs work with protein cofactors, such as Hfq or RNaseE. Interestingly, these RNAs often only require recognition of their targets by a short 15 nucleotide long sequence at their 5' ends, a mechanism reminiscent of microRNAs in the more highly evolved eukaryotes (*122*).

2.4.1.1 DsrA and RprA activating sRNAs

RpoS is an important stress response σ factor in *E. coli*. By sequentially deleting genomic sequences near the *RpoS* gene, expression of RpoS was found to be upregulated by a downstream encoded noncoding RNA, DsrA (*123*). This activation is the result of increased translation and involves the protein Hfq, an RNA-binding protein and known regulator of RpoS translation. Mutations in the 5' end of RpoS mRNA bypass the dependence for Hfq or the DsrA noncoding RNA for activation (*124*).

The structure of Hfq has ultimately revealed it to be a Sm-like protein with two RNA-binding domains. Mutation and deletion analyses suggest that Hfq binds noncoding RNAs such as DsrA with one domain and mRNAs with another to facilitate their RNA-RNA interaction (*125, 126*). Binding of Hfq and DsrA to RpoS mRNA alters the mRNA structure, relieving an inhibitory stem loop structure and promoting translation (*126*). Another *E. coli* noncoding RNA RprA was also shown to be able to bind Hfq and RpoS mRNA and activate translation in a similar fashion as DsrA (*127, 128*).

2.4.1.2 6S transcription silencing sRNA

The first bacterial sRNA, 6S, was described in 1967 (*129*) and its sequence was published in 1971 (*130*). A function for this highly expressed RNA was not suggested until recently in 2001. Gradient fractionation of this noncoding RNA reveals that it

actually migrates at 11S due to a protein cofactor. Mass spectrometry has revealed this cofactor to be RNA polymerase. Multiple analyses show interactions between 6S and bacterial RNAP. Specifically it was found that 6S preferentially binds and sequesters σ^{70} bound RNAP during the stationary phase of *E. coli* growth under nutrient limitation (*131*).

Covariance analysis between 6S noncoding RNAs from many prokaryotic species reveals a conserved structure that strongly resembles an open complex structure, a structure formed in double stranded DNA at the transcription start site of genes upon RNAP binding (*132*). This RNA structure is required for 6S transcription regulation (*133*). Thus 6S poses as an analogue to RNAP's natural DNA binding target and excludes RNAP transcription of DNA. To escape this inhibition RNAP uses 6S RNA as a template to synthesize short 14 to 20 nucleotide long pRNAs. The pRNA-6S complex no longer binds RNAP, liberating it for renewed transcription during high nutrient conditions (*134*).

2.4.2 Human 7SK RNA

7SK is a 330 nucleotide noncoding RNA expressed by RNAP III that was cloned and characterized in human Hela cells (*135-137*). A biological function for 7SK was found long after its initial discovery when it was found that the RNA forms a complex with the important transcription elongation factor P-TEFb (*138, 139*). P-TEFb is a complex comprised of the cyclin-dependent kinase Cdk9 and usually cyclin T1 but sometimes cyclin T2 or K. RNAP II elongation requires phosphorylation of the carboxyl-terminal domain (CTD) by P-TEFb. Researchers in two independent labs noted that P-TEFb purified in a small active complex and a large inactive complex. The large complex could be activated by treatment with either high salt or RNase A (*138, 139*). The RNA found complexed with P-TEFb was the 7SK RNA. It was estimated that up to 50% of P-TEFb in human Hela cells is sequestered in a complex containing the 7SK RNA (*138*).

The mechanism of 7SK regulation of P-TEFb was found in a new cofactor, HEXIM1. HEXIM1 was suggested as a binding partner for 7SK based on a yeast three

hybrid assay. *In vitro* HEXIM1 binds P-TEFb and inhibits its kinase activity but only in the presence of 7SK. Deletions studies suggested that 7SK binds a central regulatory element in HEXIM1 and releases a C-terminal domain that bound the Cdk9 subunit of P-TEFb (*140*). More recently, P-TEFb has been found to be activated by acteylation of cyclin T1 by p300. This acetylation event disassociates Cdk9 from HEXIM1 and 7SK leaving P-TEFb an active complex (*141*).

2.4.3 Human NRSF dsRNA

Neuronal restricted silencing factor / RE-1 silencing transcription factor (NRSF/REST) is a transcription repressor. Sequencing of small RNAs derived from adult hippocampal neural stem cells identified a 20 nucleotide long double stranded RNA fully complementary to the NRSF/RE-1 recognition sequence (*142*). Expression of this NRSF dsRNA increased the expression of a luciferase gene driven under a promoter containing an NRSF/RE-1 element. Expression of this dsRNA also correlated to neuronal differentiation and activation of genes contained NRSF/RE-1 elements. Electrophoretic mobility shift assay (EMSA) suggested that NRSF/REST bound NRSF dsRNA with higher affinity than for NRSF/REST dsDNA.

An attractive model for regulation would be that NRSF dsRNA competes with NRSF/REST for the NRSE/RE-1 binding element, however, ChIP studies reveal NRSF/REST still firmly bound to their NRSF/RE-1 sequences after expression of the NRSF dsRNA (*136, 142*). This leaves this observation without a biologically relevant model to explain how this dsRNA effects NRSF/REST activity. This paper overlooks the strong possibility of this effect being caused by off-target effects through the RNAi pathway, a phenomenon frequently associated with dsRNAs of this size (*70, 71*). This paper should have tested a knockdown of the RNAi pathway to show that, as the authors counter-intuitively assume, this dsRNA is not acting as a simple siRNA. Also lacking is a study of the effect of this dsRNA on transcriptionally repressed genes not regulated by NRSF/REST to show specificity. Finally, the biological relevance would have been

strengthened if the authors had performed more studies using endogenous genes rather than artificial constructs. To date no further data has been published that addresses or reproduces the data presented in this study.

2.4.4 Mammalian B2 RNA

B2 RNAs are RNAP III transcripts that are encoded with some 10,000 copies in both the mouse and rat genomes (*143, 144*). Under the cell stress of heat shock, these RNAs are strongly upregulated (*143-145*). This upregulation can be seen by semi-quantitative RT-PCR and by primer extension and other RNAP III transcripts such as 7SK are not upregulated (*145*). This upregulation of this 178 nucleotide long B2 RNA also correlates strongly with reduction of RNAP II transcription. Inhibition of RNAP II by B2 RNA can be reconstituted *in vitro* and is specific. RNAs produced by the B1 element, which is also upregulated by heat shock, and Group I intronic RNAs do not effect RNAP II transcription *in vitro* (*145, 146*).

EMSA studies suggested an interaction between radiolabeled B2 RNA and RNAP II which could be reversed by high salt or by competition with nonradiolabeled B2 RNA, however, these first studies lacked any negative controls (*146*). Finally, the same group that discovered RNAP II inhibition by B2 RNA published a more thorough study of B2 RNA structure and function. The secondary structure was suggested by RNase digestion studies using RNase V1 (cleaves double stranded RNA), RNase 1 (cleaves single stranded RNA), RNase T2 (cleaves single stranded RNA), and RNase T1 (cleaves single stranded RNA after G). Deletion analysis showed RNAP II recognition and regulation required two hairpin structures contained in 3' end of the B2 RNA. Each hairpin could bind RNAP II but could not regulate transcription, however, combination of two truncated RNAs each with one of the hairpins was sufficient to both bind RNAP II and inhibit transcription (*147*).

2.4.5 Steroid receptor RNA activator

A yeast two-hybrid screen designed to identify co-factors for the steroid hormone receptor human progesterone receptor (hPR) yielded a positive cDNA, which when placed in human cells potently activated gene expression for several steroid hormone regulated genes. This novel gene was named steroid receptor activator, SRA. However, attempts to isolate the protein produced by SRA cDNA were unsuccessful. Forcing the expression of a GST-fusion or GAL4-fused protein with the presumed protein product of SRA also failed to induce gene activation. This led to the conclusion that the SRA RNA product, an approximately 800 nucleotide long transcript, could be directly responsible for the transcriptional activation. Immunoprecipitation recovered SRA RNA associated with SRC-1 and androgen receptor, but not with p300, thyroid hormone receptor, or RXR (*148*). Furthermore, it was found that SRA specifically bound to estrogen receptor α bound SRC-1/TIF2 complex through a direct interaction with a family of RNA-binding DEAD box proteins, p72/p68 (*149*).

It was later shown that a fusion of GFP with the presumed protein product of SRA, SRAP, could activate expression of a luciferase reporter driven by the androgen receptor promoter. Deletions of the SRA cDNA and a frame shift mutation reversed the activity (*150*). Furthermore an antibody to SRAP was able to detect expression of the ~30 kD product in several human cell lines as well as skeletal muscle tissue throughout mammals and birds (*151*). However, this is no evidence of an interaction between SRAP and the SRC-1/TIF2 complex or of SRAP acting directly as a transcription factor.

On the other hand, a detailed study of the RNA structure of SRA has suggested elements of the secondary structure to be important for SRA regulation of SRC-1/TIF2 activity. Covariance analysis suggested a complex secondary structure for the 687 nucleotide long active core of the RNA that is conserved across mammalia. In this study, specific mutations designed to disrupt RNA secondary structure but preserve amino acid sequence were sufficient to reverse SRA activity. However, compensatory mutations designed to restore RNA secondary structure and either destroy or preserve amino acid sequence were not tested (*152*).

2.5 IMPRINTED NONCODING RNAS AND X-INACTIVATION

Typically every gene in the eukaryotic genome has at least two copies. These minimal two copies would be the maternal and paternally inherited copies. Presumably for the reason of controlling dosage of gene expression, some genes are only expressed from one of their genetic copies and the other copy remains silent. This process is called imprinting. Similarly, females have two copies of their X-chromosome and in this case, large sections of the extra chromosome are kept silent. However, in both of these cases, silencing is associated the expression of noncoding RNAs from the silent loci. Noncoding RNAs do not regulates imprinting in *trans*, which is sensical since it would silence both of the two identical genomic loci. However, expression of the noncoding RNAs from the one locus is only associated with silencing of that loci and not other parent's copy.

2.5.1 H19 RNA

The locus encoding the gene *Igf2* was one of the first imprinted genes discovered. Igf2 is expressed from the paternal allele. The maternal allele expresses a noncoding RNA H19 (**Fig 6a**) (*153-155*). This system became an early model of imprinting and the fact that loss of this imprinting was associated with several cancers probably contributed to the high volume of papers published since 1991. By 2003, the 251 publications accumulated, many of which contained murky and contradictory implications, inspired a review entitled "Enhancing the confusion", referring the recent work describing the role of long range enhancers in imprinting (*156*). Since then another 240 papers have been published.

The compilation of many publications concerning various deletions in the locus surrounding the *H19* and *Igf2* genes elucidated that an enhancer element located downstream of the H19 gene that is required for expression of both H19 and Igf2 (*154, 157*). This led to a model that chromosomal structural reorganization might contribute to the decision whether to express H19 or Igf2 (*154, 157*). It had also been reported that the

maternal allele is unmethylated and the paternal copy is methylated at several sites including upstream of the *H19* gene. The functional significance of this was suggested by the observation that the protein CCCTC-binding factor, CTCF binds the unmethylated copy between the *H19* and *Igf2* genes. This suggests that CTCF acts as a transcriptional insulator, shielding *Igf2* from the effect of the downstream enhancer (*158-160*). However, even after this discovery, the observation of enhancer mediated imprinting of *H19* and *Igf2* did not appear to extend to all cell systems and tissues (*161*).

The CTCF binding site is situated downstream of the H19 locus and far downstream from the Igf2 gene. This observation suggests that the H19 / Igf2 imprinting may be mediated by gene looping involving an assortment of proteins including CTCF (*162-164*). Recent literature suggests that CTCF is responsible for organizing the positioning of cohesins on the chromosome, thus rearranging intra-chromosomal interactions between distant regions of the genome (*164*). 3C analysis can detect interactions between the H19 / Igf2 loci (*162*) and *in vivo* FISH assays suggest that those interactions are lost with the knockdown of CTCF (*163*). However, the model of CTCF or cohesion mediated loops is only supported by correlations and it would be desirable to see evidence that CTCF itself mediates the loops or that the loss of the loop is what silences H19 expression or activates Igf2 expression.

Furthermore, this model does not address a very important question: what is the purpose of the noncoding RNA, H19? In addition to that, several noncoding antisense RNAs have been shown to be expressed overlapping the Igf2 coding region (*154, 156*). An attractive model might be the purpose for this noncoding transcription might be to play a role in the chromosomal organization of the locus, but more study is needed to shed light on this area.

2.5.2 Kcnq1ot1 RNA

Near the *Igf2* locus on chromosome 7 resides another imprinted locus that is regulated independently of the *Igf2* and *H19*. This locus contains seven imprinted protein coding genes and one antisense noncoding transcript contained within the gene *kcnq1*

named Kncq1ot1 (*165, 166*). Contrary to Igf2/H19, in this case the maternal copy expresses the protein coding genes and the paternal copy is silenced except for expression of the Kncq1ot1 noncoding RNA (**Fig 6b**). Like *Igf2/H19* the two copies are differentially methylated, with the maternal, unsilenced copy being methylated at several locations including upstream of the Kncq1ot1 promoter (*167*).

For Kcnq1ot1, it has been demonstrated that expression of the noncoding antisense transcript is required for silencing. This was shown by making small deletions in the promoter at important transcription factor binding sites (*168, 169*) and by inserting a premature transcription termination signal (*170*). Interestingly, while loss of Kcnq1ot1 transcription strongly reversed imprinting, the differential methylation pattern seen on the paternal and maternal alleles was preserved. However, it was noticed that one gene in the area, *Cdkn1c*, remained imprinted after truncation of *kncq1ot1 (166)*. The protein CTCF has two binding sites in the region and differentially binds the paternal allele. This has been offered as a possible explanation for the two modes of imprinting in the same region (*166, 171*).

A possibility for this regulation might be an RNAi mechanism whereby the combination of the sense and antisense transcripts could produce dsRNAs that mediate transcriptional silencing as discussed in Chapter 1. However, the problem with this model is that there must be some mechanism where the effect only works in *cis* on the correct allele (*170*). Recent ChIP-RIP studies suggest that the Kcnq1ot1 RNA remains localized to the chromatin in the 1 Mb region surrounding where it is expressed from and interacts, possibly recruiting, the polycomb silencing complexes containing proteins Ezh2 and Suz12 (*172, 173*). However, since chromatin markers and remodeling complexes are associated with the same chromosomal region expressing the RNA, perhaps it is not surprising that they can be formaldehyde crosslinked and immunoprecipitated all together. However, loss of *kcnq1ot1* does reduce levels of H3K27me3, a robust indicator of polycomb complex targeting (*172*). It still remains to be determined if the RNA molecule itself has a role in silencing or if the act of transcription is the deciding factor.

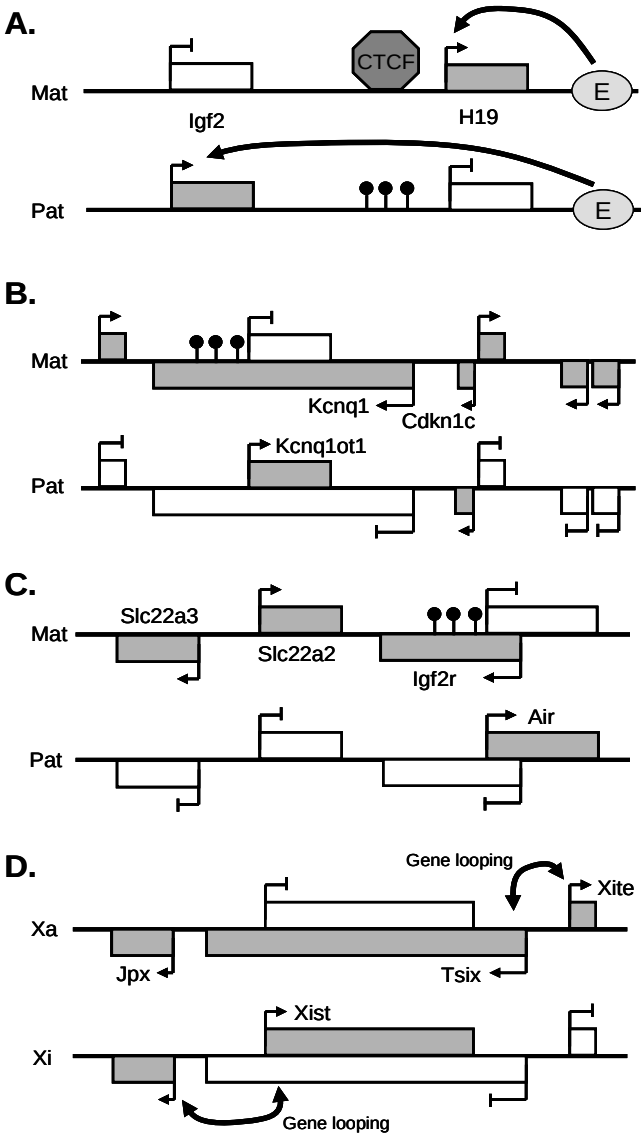


Figure 6: Imprinted noncoding RNAs and X-inactivation. (A) The noncoding RNA H19 is expressed from the maternal allele and the gene Igf2 from the paternal. Imprinting is regulated by differential methylation and recruitment of CTCF. E: denotes downstream enhancer sequence. (B) Kncq1ot1 is involved in imprinting at the Kcnq1 locus. Note that Cdkn1c escapes imprinting. (C) Air noncoding RNA regulates imprinting at the Igf2r locus. (D) Xist is expressed from the inactive X (Xi) and Tsix from the active X chromosome (Xa). This process is associated with rearrangments of gene loops.

2.5.3 Air RNA

Another example of an imprinted locus is the *Igf2r* gene locus, the maternal copy is transcriptionally active and the paternal copy is silent (**Fig 6c**). A 300 nucleotide genomic element in intron 2 of *Igf2r* is vital for imprinting and turned out to be a promoter for antisense transcript, named Air. Air is a 107,796 nucleotides long, unspliced RNA transcript (*174*). The *Air* promoter is sensitive to DNA methylation, which explains why it is expressed from the paternal allele only (*175*). The Air noncoding RNA overlaps the *Igf2r* promoter and its 3' end overlaps a distant gene *Mas1*. However, *Mas1* is not imprinted (*174*). Furthermore, two genes downstream of *Igf2r*, *Slc22a2* and *Slc22a3*, oriented in opposite directs from each other, are imprinted in placenta by Air but they are not overlapped by Air, ruling out bidirectional transcription as a model for regulation (*176*).

Evidence suggests that Air RNA 1) is required for imprinting, 2) is largely sequestered to the nucleus, and 3) interacts with chromatin. Expression of the *Air* gene is required for imprinting. Insertion of a premature termination signal shortly after the *Air* promoter abolishes imprinting (*177*). The Air RNA product is produced by RNAP II and spliced transcripts are detectable but splicing is much less efficient with only 30 to 60% of transcripts undergoing splicing. The large 108 kilonucleotide transcript is also not exported to the cytoplasm and has an unusually short half-life compared to other mRNAs (*178*). A unique RNA-TRAP assay can visualize where on the chromatin an RNA molecule associates. This is a modified FISH protocol where labeled probes are targeted to the RNA molecules of interest – in this case, the Air noncoding RNA – and antibodies conjugated to horseradish peroxidase (HRP) are directed at the hapten-labeled probes. HRP then covalently deposits biotinyl-tyramide onto the chromatin near the RNA molecule. This method revealed that Air noncoding RNA, in placental tissue where *Slc22a3* is imprinted, associated with the *Slc22a3* promoter chromatin but not in heart tissue where *Slc22a3* is not imprinted (*179*). Presumably Air might be brought in close proximity to the *Slc22a3* promoter by DNA looping.

2.5.4 Xist / Tsix RNAs

Males and females fundamentally differ by their number of X chromosomes. Females have two and males have one. X-inactivation is a mechanism to bring about dosage equivalence for many genes on the X chromosome. Most of one of the female X chromosomes is kept transcriptionally silent. The choice is random. However, from this silent chromosome is expressed a large 15,000 nucleotide noncoding RNA, Xist (**Fig 6d**) (*180, 181*). As with other imprinted loci, X-inactivation is associated with differential methylation on the one X chromosome and the other X chromosome is inactivated (*182*). Xist expression is required for X-inactivation (*183*). Furthermore, a 400 kilobase X-inactivation center including the *Xist* gene placed ectopically on chromosome 12 caused the Xist RNA to coat the entire chromosome 12, visualized by RNA-FISH, and induce chromosome-wide heterochromatin formation (*184*).

Antisense to Xist is expressed another noncoding RNA named Tsix (*185*). Tsix is expressed from the one copy and silences Xist expression to allow that X-chromosome to escape X-inactivation (*186*). Deleting the *Tsix* gene leads to ectopic X-inactivation in males and both X chromosomes to be inactivated in females, demonstrating that Tsix controls X inactivation (*187*). Upstream of *Tsix* is a small gene *Xite* that regulates Tsix expression. However, truncation of the *Xite* RNA by premature insertion of a transcriptional termination signal does not disrupt *Xite* regulation of Tsix expression (*188*). An early observation of a role for the protein CTCF (*189*) and more recently a 3C analysis of chromatin structure (*190*) suggest that *Xite* activates *Tsix* expression when they are held in a loop conformation with each other.

X-inactivation is initiated and maintained by methylation of H3K27 by the polycomb complex including the proteins Eed and Enx1 (*191, 192*). One possible role for Xist RNA is to recruit these polycomb proteins and, although only slightly suggestive, the polycomb protein Cbx7 was shown to bind RNA *in vitro (193)*. On the other hand, it was found by study of Dicer deficient cells, that Dicer-dependent small RNAs are made complementary to Xist, possibly by duplexes between the Tsix and Xist RNAs on the

active X chromosome. These duplexes keep Xist expression down. On the other chromosome in Dicer deficient cells, H3K27 methylation cannot accumulate and Xist no longer spreads to coat the inactive X chromosome (*194*). Thus RNAi might have a role in polycomb protein recruitment and spreading of X inactivation.

2.6 NONIMPRINTED TRANSCRIPTION REGULATION BY NONCODING RNAS

Even in non-imprinted genes, noncoding RNAs have been suggested to have roles in regulating gene expression. Several examples that have been demonstrated are in the early stages of characterization so the exact mechanisms and how widespread these mechanisms are used to regulate other genes remains to be characterized.

2.6.1 HOTAIR RNA

49 HOX transcription factors are clustered at 4 locations on the human genome and are potent regulators of early development and cell fate (*195*). The expression of these genes is regulated at the level of chromatin modifications. Rinn et al. performed a tiling microarray study of the 4 Hox loci (HoxA, HoxB, HoxC, and HoxD) at 5 nucleotide resolution (*195*). cDNA from 11 human tissues was annealed to the microarray revealing 231 distinguishable novel noncoding RNAs. Expression for several was verified by performing RT-PCR of the predicted RNA products. These noncoding RNAs appear to be expressed along with the other coding genes from the actively transcribed sections of the Hox loci for a particular tissue. ChIP showed that regions where noncoding RNAs and Hox genes are silenced are enriched for H3K27me3 marker and association with Suz12 from the polycomb silencing complex.

Hox loci are expressed diametrically, meaning that one region is expressed in all tissues from the anterior of the body (the head) and the adjacent region is expressed in all tissues from the posterior of the body (the feet). To more fully characterize the biological function of the noncoding RNAs, Rinn et al. chose one noncoding RNA expressed from

the HoxC locus near the boundary of two diametric chromatin domains. This noncoding RNA was named HOTAIR.

siRNA knockdown of HOTAIR did not change expression of Hox genes from the HoxC locus but instead changed expression from the HoxD locus on a different chromosome. Knockdown of HOTAIR caused upregulation of expression from the HoxD locus and reduction in H3K27 methylation and Sux12 association. Also, pulldown using a biotinylated HOTAIR RNA recovered Suz12 and EZH2 suggesting that HOTAIR associates with the PRC2 complex.

The proposed model for this regulation is that HOTAIR interacts with the PRC2 silencing complex and recruits PRC2 to the HoxD locus in *trans* by some unknown mechanism. More recently, sequencing of RNAs recovered by immunoprecipitation of the PRC2 complex have suggested association with a number of long noncoding RNAs. These RNAs are seen to be recruited to chromatin seen by RNA-FISH. siRNA knockdown of these RNAs suggest involvement in regulation of many PRC2 regulated genes (*196*). However, a specific association of HOTAIR with a particular regulated region of chromatin has not been shown.

2.6.2 CCND1 regulation

An *in vitro* screen of cell fractions for novel inhibitors of the histone acetylase CBP/p300 yielded an RNA binding protein, Translocated in Liposarcoma (TLS) (*197*). TLS has an N-terminal domain that interacts with strongly with CBP/p300 and a C-terminal domain that binds RNA. The C-terminal domain binds and sequesters the N-terminal domain until interacting with RNA, at which point the N-terminal domain is released to inhibit histone acetylation and thus transcription. This entire process can be reconstituted *in vitro* and studied using sequential addition of truncated C- and N-terminal TLS contructs and RNA and subsequent detection of histone acetylation. *In vivo*, siRNA knockdown of TLS led to robust activation of the CBP/p300 regulated gene cyclin D1 or CCND1.

RT-PCR reveals several noncoding RNAs are expressed from the promoter of CCND1. DNA damage inducing ionizing radiation, IR, is known to reduce expression of CCND1 and was found to activate expression of the promoter noncoding RNAs. Once activated, these noncoding RNAs had enhanced interactions with TLS, seen by RNA immunoprecipitation. Enhanced interactions between TLS and the CCND1 promoter were observed by chromatin immunoprecipitation. No change was seen in association of CBP/p300 with the CCND1 promoter.

Thus a model for regulation of CCND1 is that upregulation of noncoding RNAs in the CCND1 promoter recruits the RNA binding protein TLS in *cis*. Upon RNA binding, TLS undergoes an allosteric modulation releasing its N-terminal domain to bind and inhibit CBP/p300. The subsequent reduction of histone acetylation leads to silencing of CCND1 transcription.

2.6.3 p15 regulation

A small screen of genes for sense-antisense RNA transcripts yielded 111 genes that potentially have antisense transcripts overlapping their 5' ends (*198*). Among these genes was found the important tumor suppressor gene, p15, which is often deleted or hypermethylated in many cancers. 5' RACE cloning revealed the antisense transcript to be 34.8 kilobases long. In many cell lines and tissues extracted from cancer patients, there appears to be an inverse correlation between expression of p15 and the antisense transcript, p15-AS. Regulation of p15 by an antisense transcript could be reconstituted using an artificial construct with a p15 promoter followed by an antisense CMV promoter or Tet-responsive CMVTre promoter followed by GFP. GFP expression was reduced for constructs with the antisense CMV promoter or for the antisense CMVTre promoter upon the addition of Tet. This effect was unchanged even in Dicer$^{-/-}$ cells ruling out an RNAi based mechanism.

Antisense expression of p15-AS reduced H3K4 methylation and increased H3K9 methylation, signs known to be associated with gene silencing. The authors argued that H3K4 methylation was also reduced on endogenous p15 in *trans*, however, this was

difficult to see since methylation actually increased for most sites tested. Increases in H3K9 methylation for endogenous p15 was clear. Also, for mouse embryonic stem cells transfected with exogenous p15-AS, no change in DNA methylation could be detected. However, subcloning a subset of mouse embryoid bodies from stem cells revealed increase in CpG methylation seen by bisulphite pyrosequencing.

Unfortunately, this study suggests a correlation of p15 expression with expression of p15-AS but leaves no basis for formulation of a biological model for regulation. Particularly the claim that the RNA can regulate in *trans* is troublesome since the data is questionable and there is no mechanism to support this notion. This study would have been greatly helped by experiments targeted at the endogenous p15-AS rather than focusing on exogenous constructs.

2.6.4 DHFR regulation

Finally, the gene dihydrofolate reductase, DHFR, is an important cell cycle regulator that has a minor transcription start site in the sense direction in its promoter (*199, 200*). This minor start site produces a short sense noncoding transcript that overlaps the DHFR promoter and the first two exons. An *in vivo* construct containing either the major DHFR promoter, both the major and minor promoters, or the minor DHFR promoter only revealed reduced transcription seen by nuclear run-off assay for the case of both promoters oriented in the sense direction. Additionally, reduced transcription could be seen when the sense transcript was expressed in *trans*.

For a model, Martianov et al. propose two models and would appear difficult to reconcile. First, EMSA and RNA immunoprecipitation suggests that the sense noncoding RNA might interact with the transcription factor TFIIB. This suggests that the transcript may compete off TFIIB and thereby silence transcription. On the other hand, the authors suggest by EMSA that the noncoding RNA might form a triple helix with double stranded DNA.

This paper also leaves much to be desired for the demonstration of this mechanism. First, EMSA gives little support that an interaction can actually occur *in*

vivo. Second, it is difficult to understand how a competitor for TFIIB would effectively silence in *cis* without inducing widespread silencing. Lastly, this model for regulation is overly simplistic and does not account for any role of an antisense promoter that lies in between the major and minor DHFR promoters (*201, 202*).

2.7 THE TRANSCRIPTOME

The spectrum of roles that RNA fills is quite broad. With the development of modern high throughput sequencing and microarray techniques, the diversity of RNA molecules expressed from eukaryote genomes would be increased by orders of magnitude. The collection of all RNA species produced in an organism is referred to as its transcriptome. Currently, the scope of the transcriptome is just being appreciated and little or nothing is known about the function of this modern day RNA world.

2.7.1 The mammalian transcriptome

The first transcriptome to be studied with modern high throughput techniques was highly evolved mammals, rather than their simpler cousins. Using tiling microarrays with probes placed every 5 nucleotides, RNA purified from 8 different human cells lines was mapped back to the genome. Taken together from the combined studies of Gingeras and colleagues at Affymetrix Inc and the ENCODE consortium, a picture emerges that whereas less than 2% of the human genome codes for mRNA, more than 90% of the human genome is transcribed (*203-205*). The reason for this widespread transcription is unknown. Many RNA transcripts are found uniquely in the cytoplasm or uniquely in the nucleus. Some transcripts are found only in the polyadenylated fraction and others are found only in the non-polyadenylated fraction (*203*). Further characterization of the transcripts using 5' and 3' RACE revealed a complex landscape of interleaving sense and antisense transcripts with most regions of the genome showing evidence of simultaneous transcription from both DNA strands (*204*).

A similar story is seen in the mouse genome. Using a high throughput technique known as cap analysis of gene expression, CAGE, to sequence tens of thousands of transcription start sites and termination sites, widespread transcription across the mammalian genome was observed (*206, 207*). These authors described the genome as being dotted with dense "forests" of intense transcription, surrounded by long "deserts" denoted by little transcriptional activity (*206*). The authors conclude that the majority of the mammalian genome is transcribed and often from both strands. On closer analysis of pairs of overlapping sense and antisense transcripts, there is evidence of co-regulated expression. Cells were perturbed by treatment with bacterial lipopolysaccharide (LPS). Some sense/antisense pairs rose or fell in expression together. Some reciprocated with one rising and the other falling. Some moved together but after a delay as if one transcript acted as a buffer for the other. Additionally, siRNAs targeting the antisense transcript was able to induce small but reproducible changes in expression of the other, either silencing or cause small activation (*207*).

By many measures, the majority of transcription in the mammalian genome would appear to be noncoding for protein. Certainly the largest fraction of RNA transcripts lacking polyadenylation or sequestered into the nucleus could be taken as noncoding (*203*). Furthermore, analyses of the frequency of start and stop codons for these transcripts compared to known mRNAs suggest there is little possibility for coding potential (*203, 206, 208*). Many RNA transcripts are shorter truncations of a longer coding mRNA but with severely diminished or no coding capacity and thus are taken to be noncoding variants of an mRNA. These noncoding RNA transcripts have less sequence conservation than seen for coding RNAs but still show higher than average sequence conservation between mouse, human, and chicken (*206*).

On the other hand, there is a tremendous potential for coding capacity in these transcripts, especially for short proteins less than 100 amino acids long. Most protein databases assert an artificial cut off for amino acids sequences to be at least 100 amino acids long. However, analysis of novel transcripts discovered by high throughput methods suggested that many may code for short proteins that, other than their length, look very much like well known proteins in their sequence composition and conservation.

Fusion of the GFP gene to these small protein coding RNAs and their 5' UTRs did induce expression of GFP that was trafficked to various suborganelles of the cell (*209*).

More genome wide characterization of transcription as well as chromatin markers suggest a more complete picture that transcription factors, chromatin markers of active transcription, and RNA polymerase II associate with regions far beyond those that are known to code for protein (*208, 210, 211*). In fact, the vast "deserts" of low levels of transcription are found to be spanned by enormous spliced transcripts connecting protein coding regions separated by hundreds of thousands of nucleotides (*210*).

Many possible roles for noncoding RNAs have been speculated. One study found that noncoding regions of the genome that contained ultra-conserved regions (UCRs) express noncoding RNAs whose levels were altered in many leukemias and cancers. In fact, some of these transcripts are predicted to be targets for miRNAs and their expression levels change predictably with changes in the expression of the miRNA targeting them (*212*). This suggests that noncoding RNAs may have a role in tumorigenesis. Another study found that noncoding RNA levels and genome wide transcription levels are much higher in embryonic stem cells and decrease as cells undergo differentiation (*213*).

In addition to long noncoding RNAs, microarray analysis and high throughput sequencing techniques reveal a large number of short noncoding RNAs, especially near coding genes transcription start and termination sites (*205, 214-216*). These short RNAs originate from sense and antisense transcripts, with sense transcripts predominantly extending downstream of a transcription start site and antisense transcripts extending upstream. This would suggest that RNA polymerase II might bind and extend from a transcription start site in either direction. Frequently, RNAP2 is found paused near a transcription start site (*214*). This model is agreement with earlier observations that RNAP2 abundance is distributed symmetrically about transcription start sites (*210, 211*). The abundance of these transcripts is controls by the RNA exosome. Knockdown of the RNA exosome caused a marked accumulation of transcripts from gene promoters (*217*).

The role of RNA degradation may also explain the abundance of short RNAs near gene ends. A recent analysis of short RNAs with capped 5' ends revealed many to

contain splice junctions. Alignments of these 5' modified short RNAs suggests that they are degraded from mRNAs, even to contain exon boundaries (*216*). This suggests an alternative model that RNAP2 may transcribe gene loci in both directions, pausing near the gene termini accounts for the relative abundance of RNAP2 at these sites (*214*), and the RNA transcripts are degraded with those fragments overlapping the transcription start site being preserved for some unknown purpose.

Finally, both in mouse and human are found many noncoding RNAs that contain retrotransposons in them. Many retrotransposons are expressed in both directions from novel transcription start sites. Furthermore, gene 3' UTRs are generally thought of as devoid of repeat elements but many do contain repeat elements. There is a noted decrease in overall abundance of mRNA transcripts as the percentage of its 3' UTR containing retrotransposons goes up. Considering the large amount of sequence complementarity between repeat elements, this suggests the possibility of an RNA-RNA based mechanism for repeat elements to regulate expression of protein coding genes (*218*).

2.7.2 Invertebrate transcriptomes

Widespread noncoding transcription is not unique to higher order eukaryotes like mammals. A microarray based study of the *D. melanogaster* indicates that whereas around 9% of that genome codes for protein, estimates are that more than 85% of the genome is transcribed (*219*). Differing ratios of coding versus noncoding RNAs are expressed at different stages in fly larvae development. An analysis of RNAP2 localization and 10 transcription factors involved in spatial patterning in blastoderm embryos also suggests widespread transcription both in coding and noncoding regions of the genome. Many regions enriched in RNAP2 are flanked by regions enriched for the transcriptional activators. Some regions are enriched for transcriptional activators without nearby association of RNAP2. It is possible that these are examples of long distance transcriptional regulation (*220*).

In *C. elegans*, around 25% of the genome is thought to code for protein and estimates by microarray analysis suggests that more than 70% of the genome is

transcribed (*221, 222*). 454 sequencing by Shin et al. found that 50% of reads mapped to intergenic regions (*223*). Hillier et al. performed high throughput sequencing of the transcriptome during all four development stages and adult worms confirming expression of 75% of annotated genes and suggesting much of the intergenic genome is also transcribed (*222*).

2.7.3 Single celled eukaryotes transcriptomes

Genome wide studies of the fission yeast, *S. pombe*, transcriptome verifies that virtually the entire genome is transcribed (*224*). Most genes are misannotated with their 5' and 3' ends extending beyond what is annotated. Some 7000 novel noncoding transcripts have been identified. Approximately a third of those are from intergenic regions and another third are antisense transcripts. Yeast were subjected to stresses including heat stress, minimal media, and DNA damage by methyl methanesulfonate (MMS). Many noncoding RNAs modulated their expression in response to these stresses suggesting a physiological role for these transcripts. Additionally, most of the genome is transcribed in both directions. Purification of polyadenylated RNAs suggest that most antisense transcripts in *S. pombe* lack polyadenylation (*224*).

High throughput sequencing confirmed the same results with many noncoding and antisense RNAs changing their expression dramatically at various stages of meiosis or oxygen or heat stress. Additionally, sequencing revealed changes in splicing of RNA transcripts upon stress and in various stages of meiosis (*225*). A possible role for many of these noncoding RNAs is chromatin remodeling. In the case of the gene *fbp1*, transcription of an upstream noncoding RNA is required to open chromatin and allowing access for transcription factors and RNAP2 (*226, 227*).

CHAPTER THREE: Antigene RNAs

3.1 PROGESTERONE RECEPTOR

In 2004, the Corey lab began their first foray into the study of duplex RNAs targeting gene promoters. What could be appreciated from the previous literature was that promoter targeting duplex RNAs may silence transcription. The potency and sequence specificity of this phenomenon had not been demonstrated. The claims of DNA methylation were hotly contested.

The model gene used for most of this thesis progesterone receptor (PR) gene, an important prognostic marker in breast cancer. Progesterone receptor is expressed from chromosome 11 in the q22-23 locus. Progesterone receptor has a complex genomic structure with two major isoforms, PRB and PRA, expressed from alternative promoters situated some 700 nucleotides apart (**Figure 1**). The protein products differ by 165 additional amino acids on the N-terminus of the PRB protein. The two PR isoforms run on a polyacrylamide gel at around 90 and 95 kilodaltons. Unfortunately, the PR gene remains terribly misannotated in the Genbank database. Before 2008, the PRB refseq mRNA version NM_000926.3 had the transcription start site annotated to be some 720 nucleotides upstream of the published transcription start site. This has been corrected for PR refseq mRNA NM_000926.4. The shorter PRA mRNA remains unannotated in Genbank. The NCBI protein reference sequence NP_000917.3 has the longer protein misannotated as PRA.

3.2 TARGETING THE PR GENE WITH PEPTIDE NUCLEIC ACIDS

Peptide nucleic acids (PNA) are a nucleic acid mimic with a neutral and highly flexible peptide backbone rather than a phosphodiester backbone. PNAs recognize DNA through standard Watson-Crick base pairing and have an ability to invade double stranded DNA (*228-230*). These properties suggest that PNAs could be used as a tool for targeting chromosomal DNA in mammalian tissue culture.

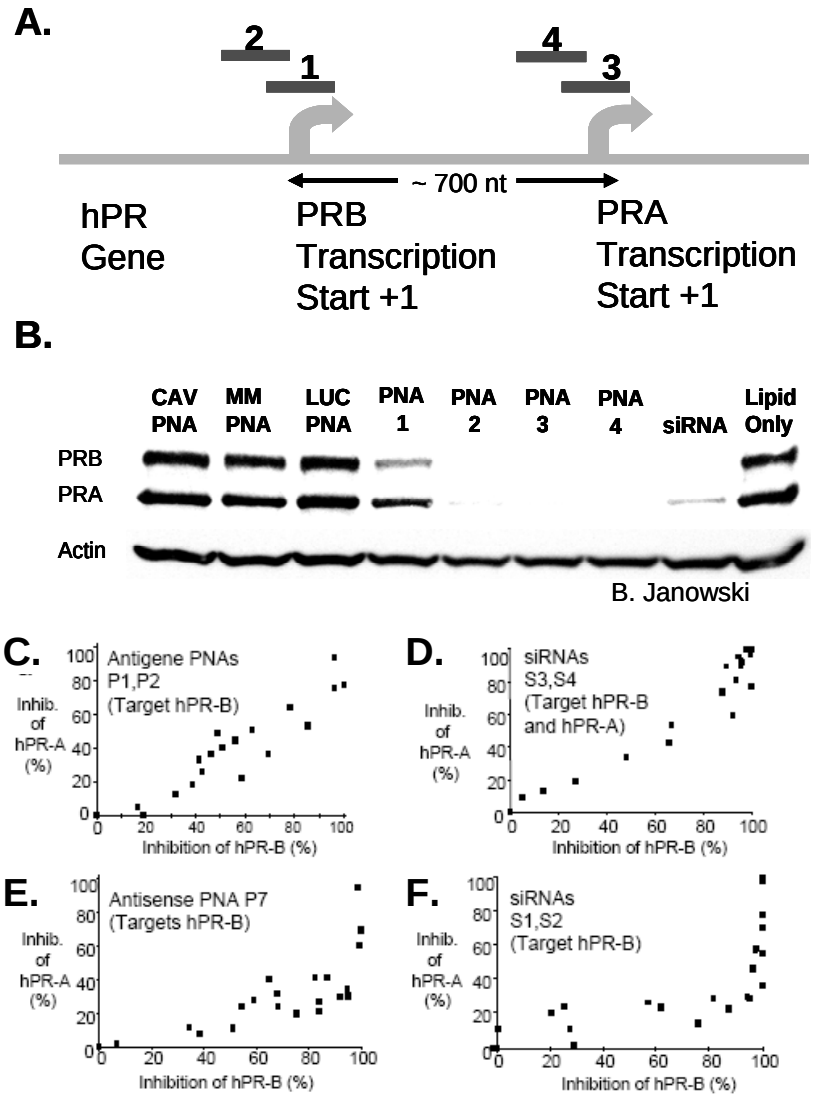


Figure 7: Peptide nucleic acids silence PR expression. (A) Four PNAs were synthesized complementary to the PR-B or PR-A transcription start sites (TSS). (B) PNAs potently silence PR expression seen by western. Negative controls target the genes caveolin or luciferase or contain mismatches. Positive control is an siRNA targeting the PR mRNA. (C) PNAs targeting the PR-B TSS potently silence both PR-B and PR-A. (D) siRNAs targeting PR-A mRNA, which is wholly contained in the PR-B mRNA, silence both isoforms. (E) Antisense PNA targeting the PR-B mRNA achieves selective knockdown. (F) siRNAs targeting the PR-B mRNA achieve selective knockdown.

B. Janowski designed four PNA sequences to target the transcription start sites for PRB and PRA (**Fig 7a**) (*231*). PNAs targeting the PR-A promoter targeted the coding strand and had no complementarity to PR-B mRNA. One mismatched PNA sequence and two other negative control sequences were also designed. I synthesized these sequences on an Expedite 8909 synthesizer (Applied Biosystems Inc.) using FMOC chemistry on resin beads. After synthesis PNAs were deblocked and cleaved from the resin with a 4:1 mixture of TFA to m-Cresol. PNAs were purified and analyzed by HPLC and fractions were analyzed by MALDI-TOF mass spectrometry to ensure synthesis success. Fractions enriched for full length PNA was then lyophilized and resuspended in water. PNAs were then annealed to their carrier DNA sequences and melting temperatures were measured using a Cary 100 Bio UV-spectrophotometer.

B. Janowski transfected PNAs into T47D breast cancer cells using Oligofectamine (Invitrogen) and potent reduction of both PR-B and PR-A isoforms was observed (**Fig 7b**) (*231*). Both isoforms were reduced whether the PR-B or PR-A promoters were targeted. This reduction was also seen at the level of mRNA. Using Scion image analysis software, I quantified knockdown for the protein levels from dose response profiles for two PNAs targeting the PR-B promoter, two siRNAs targeting both PR-B and PR-A isoforms, PNA targeting the PR-B isoform only, and siRNA targeting the PR-B isoform only. Targeting the PR-B mRNA resulted in specific knockdown until 90% of PR-B was removed, then potent reduction of PR-A was seen. However, targeting PNA to the PR-B promoter resulted in equal knockdown of both isoforms (**Fig 7 c-f**).

A model for targeting the PR promoter was to recognize the open complex structure. This structure extends from -9 to +2 with respect t to the transcription start site (*232*). This structure is known to breathe to take on a partially single stranded nature. Structures that have a partially single stranded nature are known to be susceptible to strand invasion by single stranded oligonucleotides (*233, 234*). Further research using another synthetic oligonucleotide, locked nucleic acids (LNAs), showed that other regions in the PR promoter are susceptible to targeting by single stranded oligonucleotides (*235, 236*). Pulldown experiments using biotinylated locked nucleic

acids would verify that these molecules actually sequence specifically recognize genomic DNA.

Since PNAs recognize and hybridize a sequence specific location in genomic DNA, there is a question as to why the PR-A isoform whose transcription start site is some 700 nucleotides away is potently silenced. Molecules targeting PR-B post-transcriptionally can achieve a more selective knockdown as seen by the exponentially shaped curve (**Fig 7e-f**). Several models can be proposed. One involves the proposed mechanism that RNA polymerase II scans along the DNA to find transcription start sites in order to overcome the energy barrier of finding a gene promoter in the vastness of chromatin bound DNA (*237*). For this model, a deformation in double stranded DNA at the PR-B transcription start site might be sufficient to knock RNA polymerase II from the DNA, not allowing it to find the PR-A transcription start site. However, evidence suggests that inactive single stranded oligonucleotides do recognize their chromosomal DNA targets but are unable to regulate transcription (*235*). It is difficult to reconcile the idea that one bulge in chromosomal DNA might cause RNAP2 dismissal and widespread transcriptional silencing and a nearby bulge has no effect.

The second model for PNA regulation of transcription is that the deformation cause in double stranded DNA by insertion of the PNA is detected by proteins to initiate a cellular response, whereby transcription for the region is shut down until the PNA is cleared. This model also does not explain how protein machinery would distinguish between active and inactive oligonucleotides since both appear to bind their targets.

A third model is that single stranded oligonucleotides recognize their chromosomal targets but only directly effect the transcriptional machinery within close proximity to their binding. This accounts for the observation that oligonucleotides in different orientations bind but in a less disruptive manner and are unable to silence transcription. PR-A transcription is affected as a downstream effect of the loss of PR-B transcription. Thus PR-B transcription might be a mechanism for maintaining open and transcriptionally conducive chromatin at the PR-A transcription start site. This would be a model similar to that which was seen in yeast where noncoding RNAs maintain transcriptionally active chromatin for the gene *fbp1* (*226, 227*). In support of this model,

molecules discovered later that sequence specifically active PR-B expression, also activate PR-A expression, suggesting that the model works both ways.

3.2 TARGETING THE PR GENE WITH ANTIGENE RNA

The potent knockdown of PR expression by single stranded PNAs raised the question if double stranded RNA could also target the PR promoter. B. Janowski designed three 21 nucleotide duplexes, one targeting from -2 to +17, one targeting -9 to +10, and one targeting -24 to -5 with respect to the PR-B transcription start site (**Fig 8a**). Duplexes have two T overhangs on the 3' ends. The duplex targeting -9 to +10 has 10 nucleotide overlap with PR-B mRNA and the duplex targeting -24 to -5 has no overlap with PR-B mRNA. Duplexes were transfected using Oligofectamine (Invitrogen) at 25 nM concentration. All three duplexes resulted in potent knockdown of PR expression at the level of protein (**Fig 8b**) (*238*).

To determine if effects were position sensitive, 21 duplexes targeting from -2 to -49 were tested. A well defined fingerprint was observed that one duplex potently silences PR expression and another duplex shifted over by 1 or 2 nucleotides has no effect on PR expression (**Fig 8c**). I measured melting temperatures for the 21 duplexes and quantitated knockdown by western and found no correlation between duplex stability and knockdown activity. By plotting the knockdown of the PR-A isoform vs. the knockdown of the PR-B isoform, a linear relationship is seen as was the case for PNAs targeting the PR-B promoter (**Fig 8d**).

Duplexes targeting the promoters of the genes major vault protein, androgen receptor (AR), and cox-2 also induced potent silencing of these genes. Since Morris et al. and Kawasaki et al. reported DNA methylation by promoter targeting RNA duplexes, this was tested for these RNAs. Treatment with 5-aza-cytodine and siRNA knockdown of DNMT1 had no effect on silencing by PR targeting RNAs. Furthermore, bisulfite sequencing did not detect any methylation for PR or androgen receptor promoters after treatment with duplex RNAs.

The primary concern with transfecting oligonucleotides into cells is the potential for observed effects be caused by off-target effects (*70-73*). Several aspects of this study suggest that this effect is not an off-target effect. First, are the mismatched controls. Insertion of two mismatches is sufficient to reverse activity, suggesting that this effect requires nearly full complementarity. Second, the effect is potent. For two duplexes, PR-9 and PR-13, 12 nM was sufficient to abolish almost all PR expression. Third, for multiple genes, there are multiple active sequences. Particularly, for PR there is PR-8 and PR-24 that have almost no sequence overlap and yet both still work. Fourth, the cells are healthy and grow well after transfection. Lastly, more than one technology gives the same result. Knockdown of PR induces cells to swell and adopt the same phenotype whether antisense PNA, antigene PNA, siRNA, or antigene RNA (agRNA) is used.

Several aspects of these experiments point towards protein machinery being involved. First, RNA as an isolated duplex has no mechanism to recognize its chromosomal targets. Second, DNA duplexes, which should be more stable in cells than RNA, were unable to silence PR expression. Third, inhibition by agRNAs is potent. agRNAs were able to completely abolish PR expression with one transfection at 12 nM after only three days. In comparison, agPNAs required two transfections, each at 50 nM, and 8 days before PR expression was completely abolished. Lastly, the sensitivity of position of the target where one duplex works and a duplex shifted over by one or two nucleotides does not, suggests protein involvement.

A striking feature of agRNA activity is the fact that one duplex potently silences gene expression, and a duplex moved over by one or two nucleotides has no activity. Several possibilities could account for this. One is that inactive duplexes do not hybridize strongly with their chromosomal targets. However, melting temperature data suggests that this is not the case. Second, there is the possibility that the active duplexes follow a periodicity such as the 11 nucleotides of an A-DNA or RNA helix turn. It is difficult to rule this out as a contributing factor but it does not explain why -9 and -13 work but -11 has no activity. Another possibility is that the protein machinery selects which duplexes are loaded. Lastly, there is the possibility that activity requires the protein machinery to be positioned just properly to make contacts with proteins on the DNA.

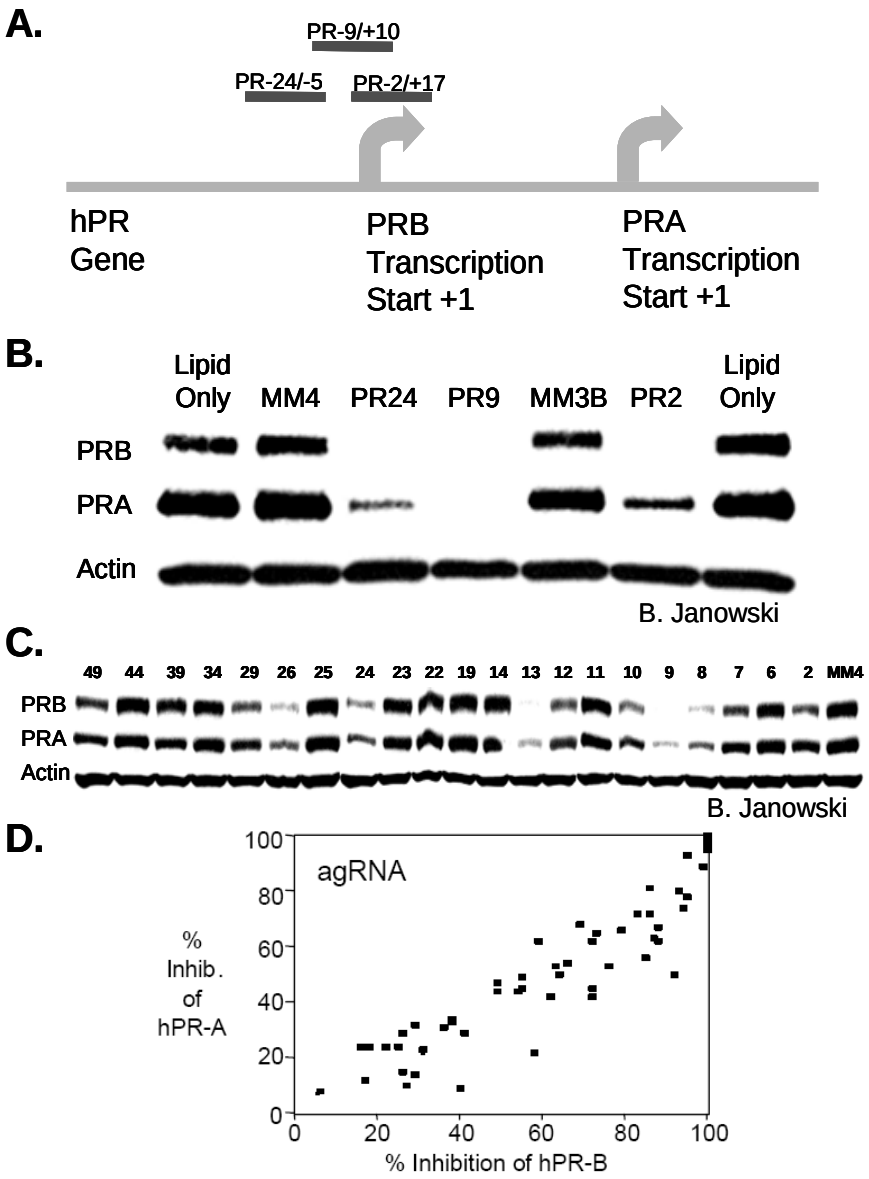


Figure 8: Antigene RNAs silence PR expression. (A) Duplex RNAs were designed to target the PR-B promoter. (B) Western blot shows potent silencing of PR expression at 25 nM concentration. (C) A pattern is seen with some duplexes silencing PR expression and nearby duplexes being inactive. Trnasfections are at 25 nM and positions are given as nucleotides upstream of the PR transcription start site. (D) As with agPNA, agRNA induces potent knockdown of both PR-B and PR-A isoforms.

Finally, as was seen with antigene PNAs, agRNAs cause linear knockdown for both PR-B and PR-A isoforms. This feature could be a cascade effect from loss of PR-B transcription as may be the case for PNAs and LNAs. An additional possibility is a mechanism exists for agRNA activity to spread to adjacent regions. This is seen for *S. pombe* where recruitment of double stranded RNA induced the production of more double stranded RNAs from adjacent regions by a direct interaction with an RNA-dependent RNA polymerase (RDRP) (*43*). Additionally, the noncoding RNA Xist has an ability to spread silencing but the mechanism remains obscure (*183, 184*). Humans have just recently been shown to possess an enzyme that may be able to act as an RDRP (*239*). Alternatively, a model could be proposed that human transcriptional gene silencing may be seeded by targeting with a duplex RNA and spreading to adjacent regions may be mediated by protein recruitment without the need for further double stranded RNA production.

3.3 AGO2 MEDIATES agRNA TRANSCRIPTIONAL GENE SILENCING

In order to distinguish promoter targeting duplex RNAs from siRNAs that target the mRNA, we refer to RNAs whose target is chromosomal DNA as antigene RNAs (agRNAs). agRNAs are identical to siRNAs in every way except that they do not target the mRNA. Another key difference between siRNAs and agRNAs is that agRNAs silence transcription and siRNAs do not. This can be seen by nuclear run-on assay where chromosomal bound RNA polymerases are allowed to transcribe with ^{32}P labeled nucleotides to directly measure transcription (**Fig 9a**) (*240*).

Because of the strong structural similarities between agRNAs and siRNAs, it is reasonable to hypothesize that the RNAi machinery could be involved in their activity. To test this hypothesis, B. Janowski designed siRNAs targeting the two major argonaute proteins in human, Ago1 and Ago2. I measured knockdown of Ago1 and Ago2 mRNA by quantitative real time PCR (**Fig 9b**). Upon demonstrating potent knockdown of Ago1 and Ago2, B. Janowski performed sequential transfections to test if these proteins are required for agRNA activity. First, siRNA targeting either Ago2 or Ago1 was transfected.

Three days later, agRNAs targeting either PR, huntingtin (HTT), or AR were transfected. Following harvest, protein expression was measured by western. In the absence of either Ago2 or Ago1, agRNA activity was reversed for 8 active duplexes targeting PR, HTT, and AR (**Fig 9c**). Additionally, Ago2 and Ago1 were able to be immunoprecipitated with the promoter DNA for PR with chromatin immunoprecipitation.

However, despite this convincing evidence, it is reasonable that although both Ago1 and Ago2 might be involved to some level, they are likely not both required. The reason is that if Ago2 were knocked down, Ago1 should be able to compensate or vice versa. By this logic, perhaps for a less potent agRNA, loss of Ago1 might reduce agRNA activity enough to appear as reversal but for a more potent agRNA or a more potent transfection, Ago2 might be sufficient to induce TGS alone. Recent data in the Corey lab suggest that for the most potent agRNAs, Ago2 alone may be all that is required (Y. Chu unpublished results).

Typically, RNAi is discussed in the context of its role in the cytoplasm (*241, 242*). Although the RNAi community took this as a surprise, this study was not the first and would not be the last to suggest a role for argonaute in the nucleus. Earlier work suggested that RNAi could reduce the abundance of the nuclear RNA 7SK (*243*). RISC isolated from the nucleus was shown to be able to cleave its cognate target for four different transfected siRNAs. Similarly, fluorescently labeled siRNAs with a nuclear target were found to be efficiently sequestered to the nucleus (*244*). Immunoflourescent studies using antibodies against the human Argonaute proteins came to different conclusions that argonaute was in the nucleus or was completely excluded from the nucleus (*245, 246*). A problem with these types of studies is with interpreting the intensity of staining for argonaute, with one image fluorescing so brightly the entire cell appears permeated with argonaute proteins and another image is dimly fluorescent so that only the most punctate concentrations of argonaute protein are visible.

Finally in *C. elegans* a systematic genetic study of nuclear RNAi uncovered an essential nuclear localization signal (NLS) carrying argonaute protein, *nrde-3*, required for nuclear RNAi. Interestingly, this argonaute is not catalytically active. Thus it is suggested that NRDE-3 may carry small RNAs into the nucleus and some hand-off

mechanism allows NRDE-3 to transfer its guide strand to another argonaute protein in the nucleus for the catalytic destruction of RNA targets (*241, 247*). However, the necessity for catalysis in nuclear RNAi appears to be assumed. Recently, two argonaute proteins WAGO1 and CSR-1 have been shown to be involved in chromosomal targeting and silencing of transposons and chromosomal segregation respectively (*248-250*).

More recently in human, nuclear RISC has been purified and distinguished from cytosolic RISC. Fluorescence correlation spectroscopy estimates using EGFP-fused Ago2 are that nuclear RISC has a molecular weight of 158 kDa and cytosolic RISC has a weight of 3MDa. Separation of nuclear RISC and cytosolic RISC over a sucrose density gradient reveal that most cytosolic human Ago2 fractionates in a protein complex larger than 350 kDa (*251*). The larger size of cytosolic Ago2:RNP is likely due to association with other large RNPs such as the ribosome (*252, 253*). Finally, it has recently been shown that shuttling of Ago2 in and out of the nucleus is mediated by Importin8 (*254*).

For the B. Janowski study, cells were treated with the histone deacetylase inhibitor trichostatin A (TSA) and this was unable to reverse gene silencing (*240*). Also chromatin immunoprecipitation suggested no changes in H3K4 dimethylation or H3K9 dimethylation. The TSA result is clear and straightforward. Perhaps for other genes, histone deacetylases may have a role to play but they are not required. For the chromatin modifications, these results are more ambiguous. The normalization for this experiment is awkward because GAPDH is shown on the side rather than normalized against. Also the error bars look small like technical replicates rather than biological replicates and the error for the negative controls is not shown. I have recently done ChIP of H3K9me3 and see only mediocre enrichment (Chap 7). On the other hand, I have also recently found large changes in H3K27me3 induced by silencing agRNAs (Chap 5).

In contrast, other labs report a correlation of chromatin modifications with RNA-directed transcriptional silencing (*69, 255-260*). Only two studies report notable changes in H3K9me2 on the order of 10-fold or more (*255, 256*). Other reported changes are quite marginal. Another recent study reports no changes in chromatin modifications (*261*). The desire to see changes in chromatin is strong. Chromatin modifications serve as a cellular memory passing on to future generations messages about which genes are to be active

and which remain silent (*262, 263*). The hope is that changes in chromatin modifications might prove to make agRNAs more robust and longer lasting than standard RNAi knockdown strategies (*258, 264, 265*). However, the only reports claiming long lasting antigene effects are marginal, with mediocre knockdown, and poorly controlled, and for these reasons they are sequestered to more obscure journals (*258, 265, 266*). If as most studies suggest, agRNA silencing is temporary and reversible (*240, 261*), then occasionally observed chromatin modifications must be a side effect to changes in transcription and not associated with programmed changes in cellular memory. It is therefore not reasonable to presume that the primary mechanism by which agRNAs silence is through epigenetic modifications.

B. Janowski's tested silencing agRNAs in a cell line that lowly expresses PR gene, MCF7 breast cancer cells. In this cell line, agRNAs are unable to reduce the already low levels of PR expression but siRNAs are still potent. This result further suggests a difference between the mechanism of agRNAs and siRNAs. However, it also suggests that something about the state that the promoter is in decides whether it will be susceptible to TGS. Genes with high levels of transcription are susceptible to silencing. Genes with low levels are not. In this context, chromatin states may have a role in TGS. Chromatin may have a role in deciding how accepting a gene locus is to regulation by agRNAs. However, assumed interactions between agRNA:argonaute complexes with chromatin remodeling enzymes based on modifications than sometimes associate and sometimes do not associate with silencing (*256, 260, 265, 267*) are presumptuous.

At the same time that the Janowski study was published, another study was published claiming the Ago1 alone was necessary and sufficient for agRNA induced TGS (*256*). The model gene in this study was a CCR5-GFP fusion. From experiment to experiment, knockdown of CCR5-GFP by a single active agRNA ranged from 40% to 70%. Knockdown of Ago1 reversed CCR5-GFP silencing for one active duplex. ChIP of Ago1 associated with the promoter DNA of the CCR5-GFP construct. Ago2 was not able to be recovered by ChIP and no follow up investigation for the role of Ago2, such as RNAi knockdown of Ago2, was pursued.

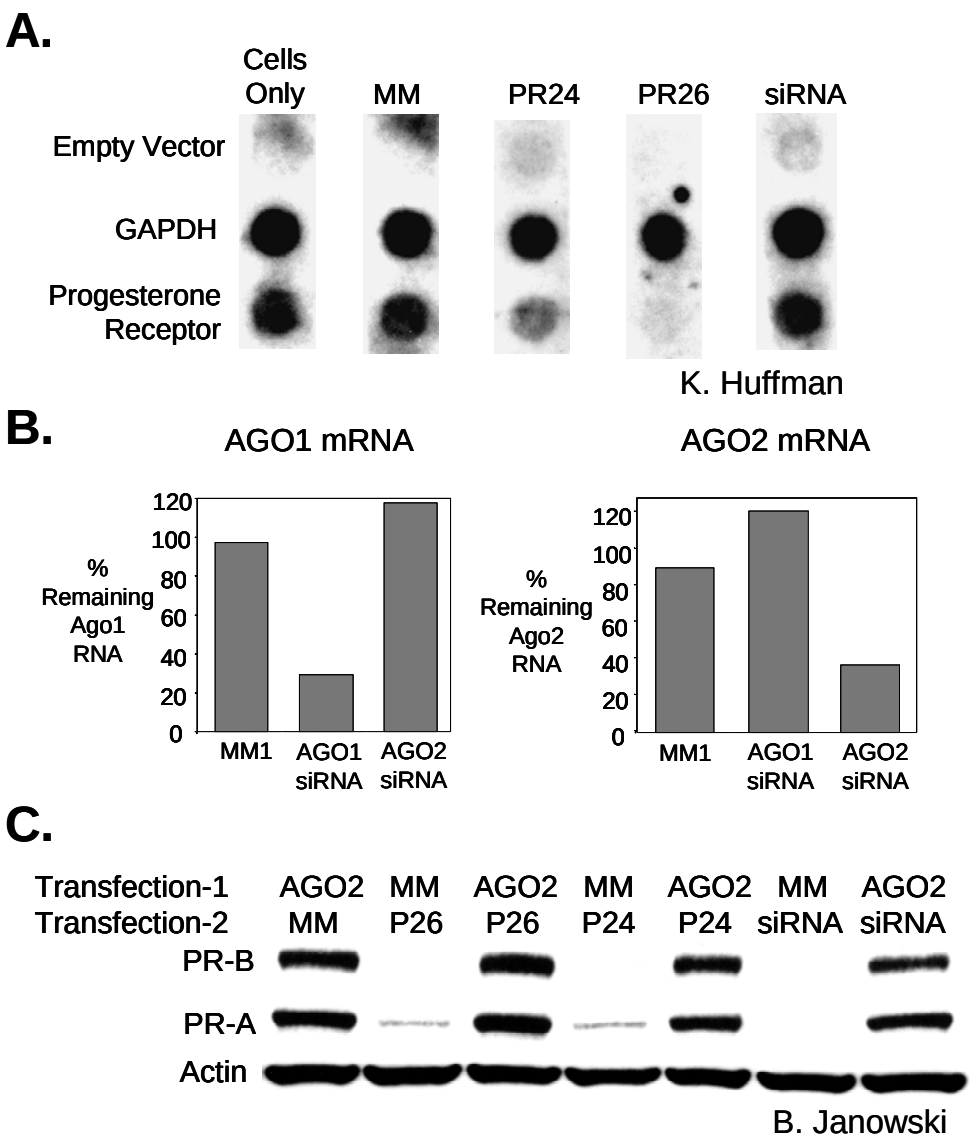


Figure 9: Ago2 is involved in agRNA TGS. (A) Nuclear run-on assay shows that unlike siRNAs, agRNAs silence at the level of transcription. (B) siRNAs were transfected and screened by real time quantitative PCR to identify duplexes that potently knockdown Ago1 and Ago2. (C) Knockdown of Ago2 reverses activity of agRNAs as well as siRNAs as seen by western.

3.2 agRNA INDUCED GENE ACTIVATION

In the case of agRNA silencing, many duplexes targeting the PR gene promoter were inactive (*238*). It was noticed that small but reproducible levels of PR protein activation could be seen for duplex RNAs that were thought to be inactive. B. Janowski developed a hypothesis that gene activation might occur but may be masked by the high levels of PR expression in T47D breast cancer cells.

To test this hypothesis, B. Janowski tested the effect of these "inactive" agRNAs in another breast cancer cell line, MCF7, which constitutively expresses low levels of PR (**Fig 10b**). In this cell line, agRNAs that were thought to be "inactive" robustly activated PR expression both at the level of protein and mRNA (**Fig 10c**). In T47D cells, hydrophobic molecules in the serum have a role in maintaining the high levels of PR expression. B. Janowski reduced the levels of PR by growing T47D cells in media stripped of hydrophobic molecules by charcoal filtering. Under this condition, "inactive" agRNA duplexes were transfected resulting in robust activation of PR expression at the level of protein and mRNA (**Fig 10a**) (*268*).

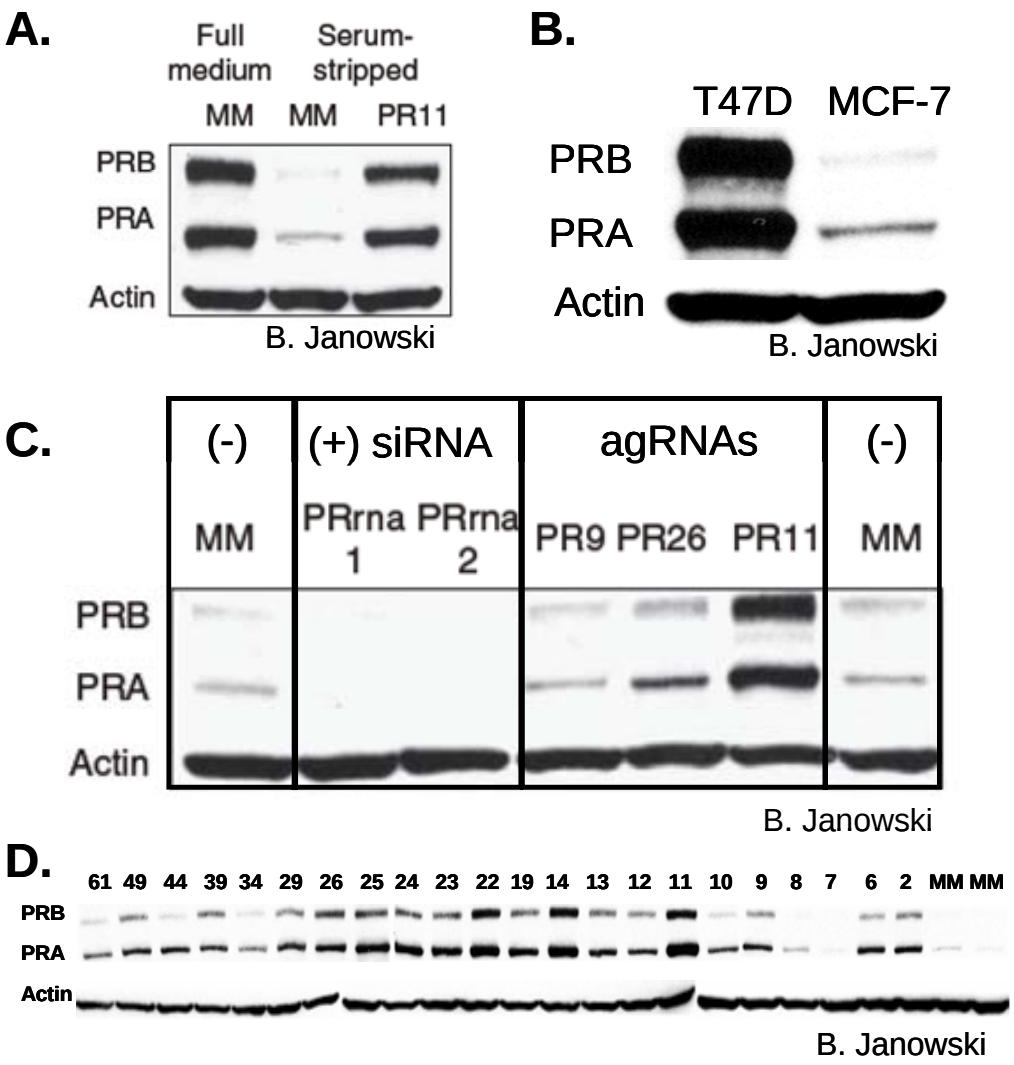


Figure 10: agRNAs can activate gene expression. (A) The duplex PR11 induces gene activation in T47D cells under serum stripped conditions. (B) MCF7 cells constitutively expression lower levels of PR than T47D seen by western. (C) The duplex PR11 can induce potent gene activation in MCF7 cells. (D) As with TGS, a pattern of active and inactive duplexes is seen for activating agRNAs. Positions are given with respect to the transcription start site for PR.

Gene activation is dose dependent. Gene activation was also reversible with low levels of PR expression being restored some 10 days after transfection. Gene activation by agRNAs was associated with changes in chromatin modifications typically associated with activation for the PR gene in this cell line, including H3K4 methylation and histone acetylation. Also, agRNA duplexes were found that activate the gene major vault protein, MVP (*268*).

As with gene silencing, a pattern appeared that some duplexes activated gene expression and immediately adjacent duplexes did not effect PR expression in MCF7 cells (**Fig 10d**). B. Janowski designed an experiment to address whether inactive RNA duplexes were unable to reach their targets or if they did bind but were unable to effect transcription. To address this she designed a competition experiment between an active duplex, PR-11 targeting from -11 to +8 with respect to the PR TSS, and an inactive duplex overlapping the target, PR-8 targeting from -8 to +11 (**Fig 11a**).

When these duplexes were transfected sequentially with PR-11 and MM or PR-8 and MM, the same effect was seen as with single transfections. PR-11 activated PR expression and PR-8 did not effect PR expression (**Fig 11b**). However, if PR-11 was transfected after PR-8 transfection, then PR-11 was unable to activate PR expression. Similarly, the inactive duplex PR-12 could compete for the target site and prevent PR-11 from activating PR expression. This meant that inactive duplexes could bind their targets but that somehow, likely due to improper orientation on the promoter to make contact with other protein co-factors, they are unable to effect transcription.

One possible objection to the competition experiment is that the first duplex transfected saturates the RNAi machinery and thereby preventing the active PR-11 duplex from activity. However, this explanation is unlikely since transfections with PR-11 following MM transfections were just as potent as PR-11 transfections alone. That said, it would have been desirable for the experiments to have been done at a lower concentration than 100 nM.

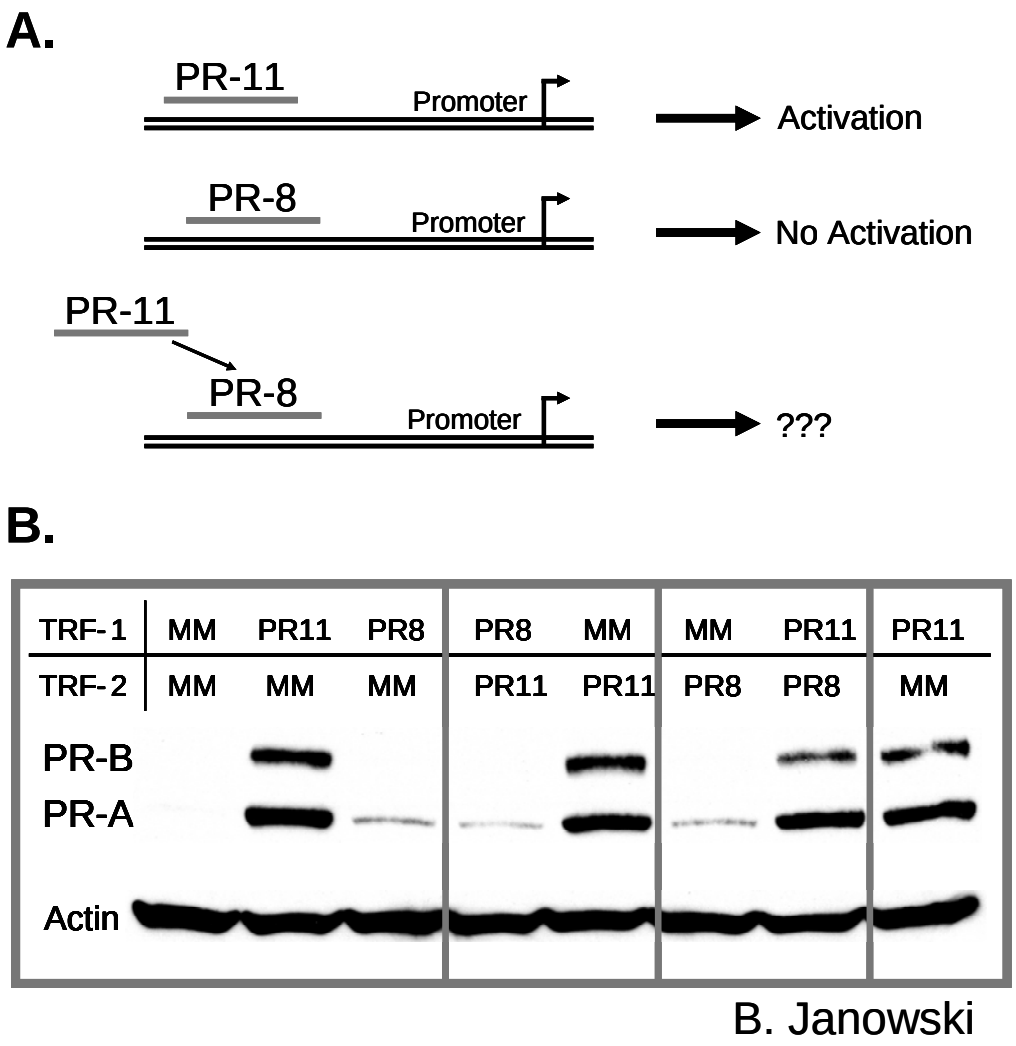


Figure 11: Competition between active and inactive duplexes. (A) Experimental design for sequential transfections of active and inactive duplexes. (B) The inactive duplex PR-8 ihibits PR-11 from activating PR expression, suggesting that inactive agRNA duplexes are able to recognize their chromosomal targets.

RNA induced gene activation was also accompanied by changes in chromatin modifications. An increase in H3K4 methylation was seen by chromatin immunoprecipitation. Also a decrease in histone H3 acetylation was seen. The histone deacetylase inhibitor, trichostatin A (TSA), and the methyltransferase inhibitor 5'-deoxy-5'-(methylthio)adenosine (dMTA) both prevented agRNA induced gene activation.

For this unexpected observation of gene activation, it is possible that the effect may be an off target effect. There are several reasons to believe it is not. First, PR-11 (targeting from -11 to +8) and PR-22 (targeting -22 to -3) both potently activate but have little sequence similarity and no potential for similar off-target effects. Second, mismatches introduced, even those clustered at one end of the duplex to preserve seed sequence complementarity for bases 2 through 7, abolish activity, suggesting the result is sequence specific. Third, there are many active duplexes given. Fourth, expression of a potent PR regulator, ERα, was monitered and remained unchanged. Lastly, activation could be induced in two cell lines, MCF7 and T47D.

For this paper, as with TGS, PR-A is upregulated with PR-B. Also treatment with TSA suggests involvement of a histone deacetylase. Histone deacetylases are known to be potent transcription regulators that can have effects over long distances (*197*). As for TGS in *S. pombe*, perhaps transcriptional gene activation has a mechanism for spreading (*43*). On the other hand, activation of PR-B transcription may induce changes in the PR-A promoter that make it more transcriptionally active, as has also been reported in yeast (*226, 227*).

Also for this paper is seen the pattern of some duplexes work and others do not. Again, duplex stability is an unlikely explanation. Also, due to the competition experiment between PR-8, PR-12, and PR-11 suggests that the decision is not made at the level of loading the duplex into argonaute protein. Inactive duplexes reach their chromosomal targets as well. Instead, the decision must be made at the level of chromatin. Some duplexes, when they recognize their chromosomal targets, are not positioned properly to interact, recruit, or block other proteins that are required to regulate transcription.

Shortly before this paper was published, another paper appeared demonstrating RNA activation by agRNAs, referred to in this paper as RNAa (*269*). This paper showed activation of three genes in three cell lines. This paper showed that activation was dose dependent. This paper also concluded that Ago2 was required for gene activation.

The Li et al. paper came to difficulty in experiments designed to show sequence specificity. There 5 mismatches were placed on either the 5' or 3' ends of duplexes. The duplexes with 5 mismatches on the 5' end failed to activate gene expression but the duplexes with 5 mismatches on the 3' end activated as well or better than a fully complementary duplex. This lack of sequence specificity was shown for two duplexes activating two different genes.

CHAPTER FOUR: Mechanism of agRNA silencing and activation

4.1 MODELS FOR agRNA RECOGNITION OF GENE PROMOTERS

The fact that argonaute proteins are involved in the agRNA mechanism still leaves their molecular target unclear. Two models can be proposed. The simplest model for agRNA targeting would be for the RNA to base pair directly with the genomic DNA (Model 1, **Fig 12**). Then argonaute can either directly or through the aide of co-factors relay the signal for POL II to down-regulate or up-regulate transcription. On the other hand, argonaute proteins are known to mediate RNA-RNA interactions. Thus there is the possibility that there exist some undiscovered RNA transcript expressed from the promoters of genes that could serve as a substrate for agRNA recognition. As a nascent transcript, this hypothetical RNA could tether the argonaute-agRNA complex to the DNA bringing argonaute and its co-factors in close proximity to chromosomal DNA much in the same way shown for the RITS complex in *s. pombe* (Model 2, **Fig 12**).

Even with the involvement of argonaute proteins, direct recognition of DNA by RNA is not unreasonable. The recognition of genomic DNA by RNA to form R-loops has been hypothesized in several other mechanisms (*199, 270, 271*). Additionally, the targeting of RNA molecules by DNA oligonucleotides to form RNA-DNA hybrids for degradation by the endogenous enzyme RNaseH is a knockdown strategy widely used in antisense oligonucleotide research (*14*). Finally, the argonaute protein from the bacteria *a. aeolicus*, whose crystal structure has been solved, mediates RNA-DNA interactions by using a 21 nucleotide long DNA guide strand for recognition of mRNA targets (*272*).

As seen in chapter 2, the human genome has many overlapping RNA transcripts in both directions. On the other hand, gene promoters are very important for regulating gene expression and many protein transcription factors have to recognize the promoter for transcription to occur. How a promoter would balance between recruiting the necessary protein machinery to activate transcription with polymerases occasionally passing through and knocking all that machinery off is complicated to imagine.

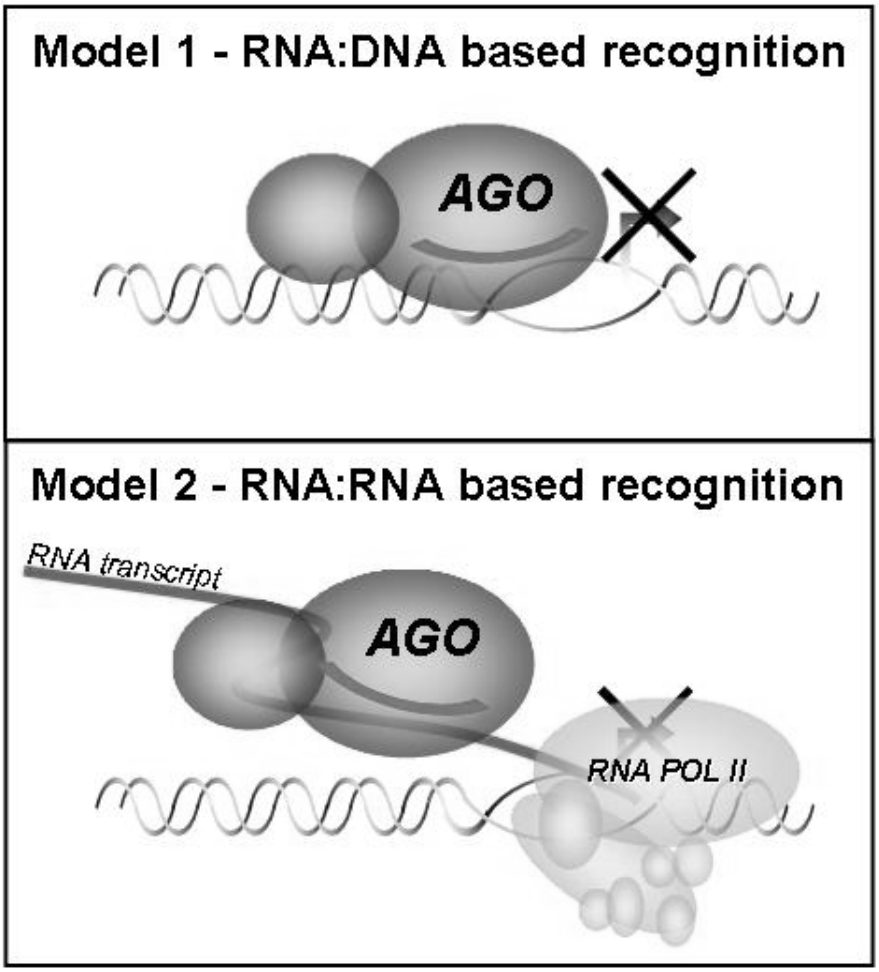


Figure 12: Two models for agRNA recognition of gene promoters. agRNAs may recognize promoter DNA directly through the formation of an RNA:DNA hybrid and thereby either directly interfere with transcription or recruit co-factors that effect transcription (Model 1). Alternatively, agRNAs may bind a noncoding transcript expressed from the gene promoter and then recruit co-factors that effect transcription (Model 2).

4.2 TRANSCRIPTION IN THE PR PROMOTER

To test if an RNA transcript exists in the PR promoter, which would be required for Model 2, I used reverse transcriptase polymerase chain reaction (PCR) amplification (*273-275*). I purified total RNA from T47D cells using a spin column protocol, GenElute® Mammalian Total RNA miniprep kit (Sigma-Aldrich). I designed PCR primers so that one reverse primer remained fixed and the forward primers moved successively upstream to produce longer PCR product (**Fig 13a**). Some primers were downstream of the PR transcription start site and some bracketed the transcription start site spanning into the promoter.

It would have been desirable to have the reverse primer in another exon to prevent detection of genomic DNA, however, exon 1 for the PR gene is more than 2300 nucleotides long. Several attempts to PCR amplify the whole exon 1 with varied Mg^{2+} and DMSO concentrations were unsuccessful, likely due to secondary structure in GC-rich tracks within the exon. Use of a proofreading Taq polymerase (Platinum Taq®, Invitrogen) was also not able to amplify exon 1. For a positive control, the 4500 nucleotide mRNA transcript for the p53 gene could easily be amplified.

Since reverse primers could not be placed in exon 2, I designed a reverse primer 250 nucleotides downstream of the transcription start site. This would mean that the primers could detect genomic DNA, which might contaminate the RNA sample. PCR of RNA samples without reverse transcriptase treatment could easily detect contaminating genomic DNA. To remove this, I treated RNA samples with 2U of DNase I (Invitrogen) for 10 minutes at room temperature and heat inactivated at 75 ºC for 10 minutes. Even after two sequential DNase I treatments, small levels of contaminating genomic DNA were still detectable after 40 cycles of PCR.

I then treated RNA samples with restriction enzymes MslI and AvaI for 1 hour and heat inactivated at 75 ºC for 15 minutes. These restriction enzymes cut at sites +91 and +107 within the genomic DNA so that PCR primers could not amplify genomic

DNA. Since the restriction enzymes only cut double stranded DNA, samples already treated with reverse transcriptase could be treated with restriction enzymes without harming the cDNA. After treatment with DNase I and restriction enzyme digest, no contaminating genomic DNA was detectable by PCR (**Fig 13b**).

Even though genomic DNA could not be detected by PCR, after reverse transcriptase treatment of 2 μg of RNA at 37ºC for 2 hours using random primers followed by heat inactivation at 75 ºC for 10 minutes (Applied Biosystems Inc.), PCR product is detected in PCR reactions using 33 ng of cDNA template (**Fig 13b**). This PCR product is from an RNA transcript that overlaps the PR promoter. The RNA transcript extends at least beyond 150 nucleotides upstream of the transcription start site for PR.

cDNA is single stranded DNA but PCR uses forward and reverse primers so it cannot detect which strand the cDNA copy is from. The upstream transcript detected may be from a rarely used upstream transcription start site for the PR gene. Alternatively, the transcript may be transcribed from the nontemplate strand, antisense to the PR gene. To distinguish whether the transcript was oriented in the sense or antisense direction with respect to the PR gene, I performed reverse transcriptase (RT) treatment with specific primers rather than random primers (*276, 277*). After treatment with primers in either the sense or antisense direction, I performed PCR to detect the cDNA product (*278*).

RT treatment with primer RE+250 to detect a sense transcript allowed detection of the PR mRNA with primers downstream of the transcription start site (+33) of PR. However, no sense transcript from the PR promoter was detected. RT with primers in the promoter to detect an antisense transcript did detect an RNA transcript suggesting that the promoter transcript may be antisense to the PR mRNA (**Fig 13c**).

However, for each RT reaction, primers directed at 6 other genes were used for negative controls to test for nonspecific amplification. In every RT reaction, at least one out of the 6 nonspecific control genes would amplify suggesting that nonspecific amplification may be a problem with this protocol. Today this protocol, referred to as strand specific PCR (*278*), is widely used to distinguish sense from antisense transcripts (*261, 265*) without regard for the possibility of nonspecific detection of transcripts.

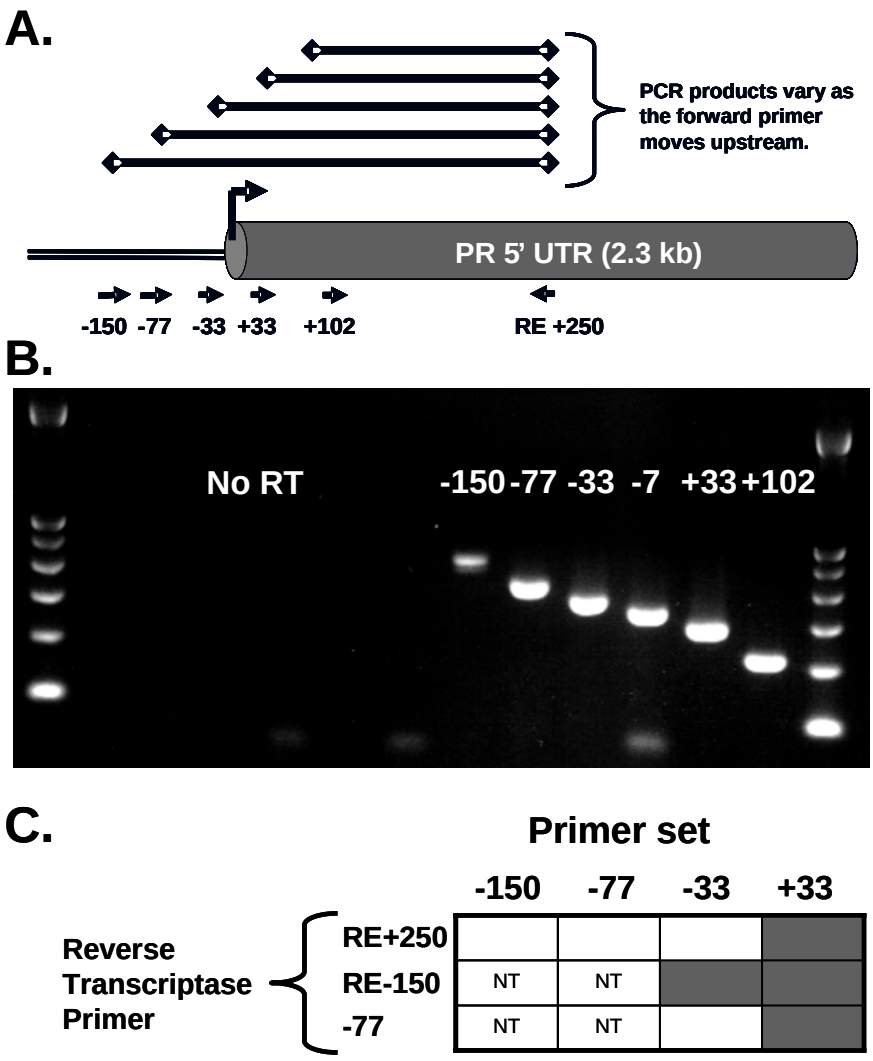


Figure 13: An RNA transcript is expressed from the PR promoter. (A) PCR primers were designed so that the reverse primer remained fixed (RE+250) and forward primers moved upstream into the PR promoter. (B) An RNA transcript is detected in the PR promoter in T47D cells as far upstream as -150. (C) Strand specific RT treatment detected the mRNA and an antisense transcript in the PR promoter. Shaded squares indicate PCR product detected. White squares indicate no PCR product detected. NT means those primers were not tested for that sample.

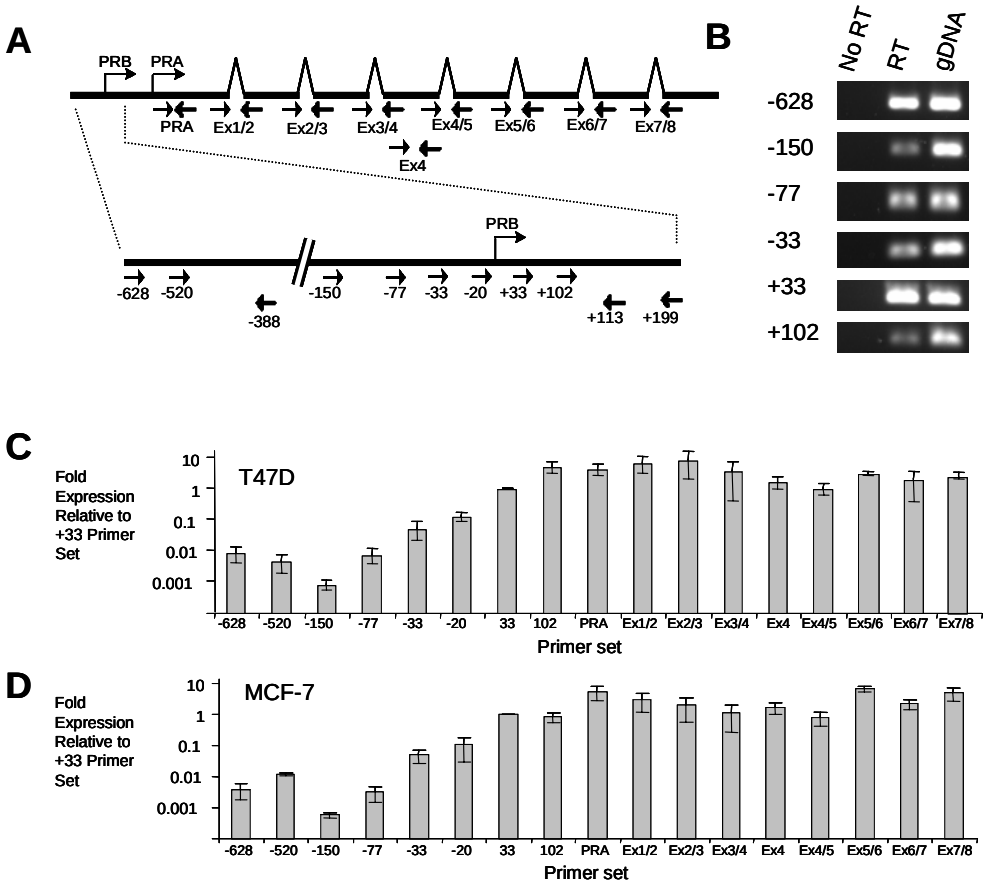


Figure 14: RT-PCR detection of an RNA transcript. (A) Schematic of primers used to detect RNA transcripts within the PR mRNA and in the PR promoter. (B) RT-PCR of cDNA prepared from T47D polyA purified RNA suggests the PR promoter RNA is polyadenylated. (C) qRT-PCR of cDNA prepared from polyA purified RNA from T47D cells reveal the promoter RNAs range from 10 to 1000 fold lower expressed than PR mRNA. (D) qRT-PCR of cDNA prepared from polyA purified RNA from MCF7 cells reveal a similar pattern as seen in T47D cells. Error bars are calculated as standard deviation and n = 6 for each cell line.

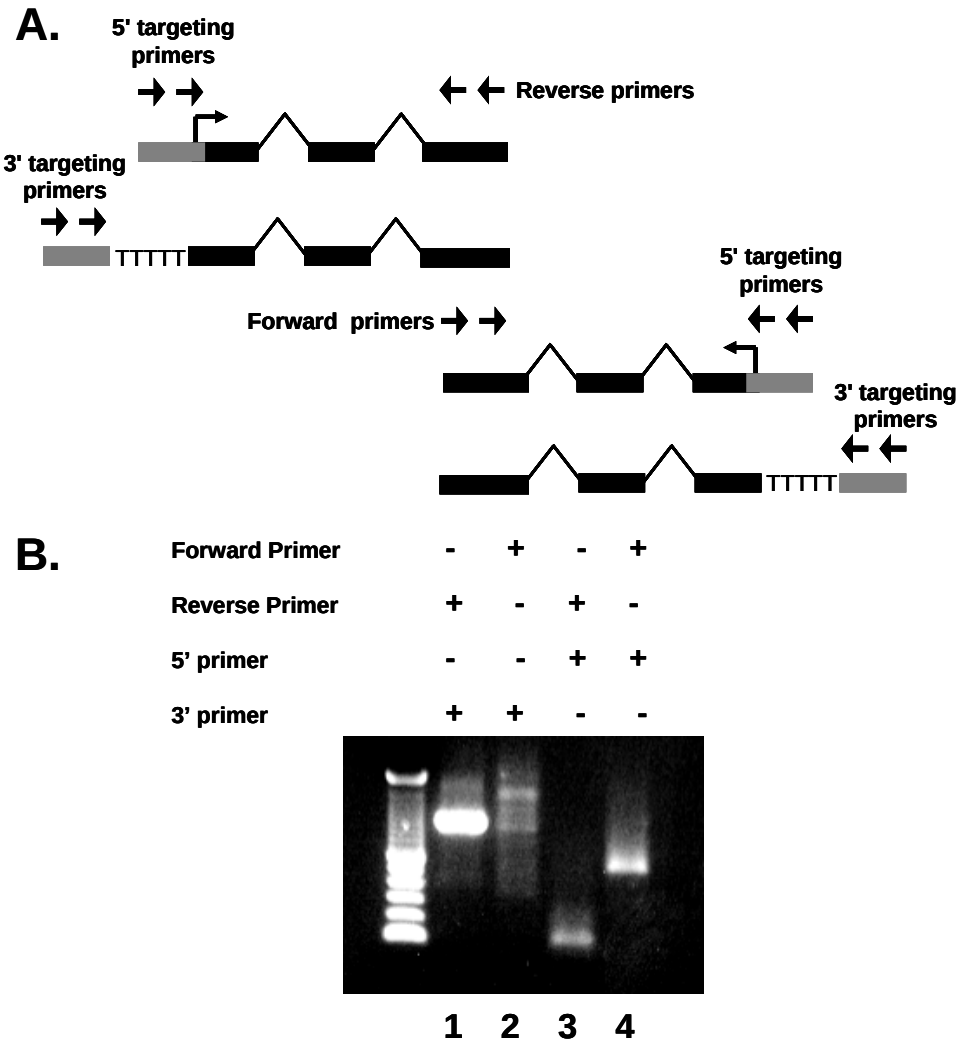


Figure 15: 5' and 3' rapid amplification of cDNA ends (RACE). (A) Using primers targeting either a 5' or 3' linker attached to cDNA ends and gene specific primers in the region of interest, cDNA ends of sense or antisense transcripts can be PCR amplified. (B) PCR product is visualized by gel electrophoresis. Products were cloned and identified by sequencing: Lane 1, 3' end of an antisense transcript overlapping the PR promoter; Lane 2, nonspecific product; Lane 3, the 5' end of PR mRNA; Lane 4, the 5' end of an antisense transcript overlapping the PR promoter.

Next I tested if the promoter RNA transcripts were stable polyadenylated RNA transcripts. After RT treatment of polyA RNA with random primers, cDNA transcripts were detected with primer sets extending up to more than 600 nucleotides upstream of the PR transcription start site suggesting that the promoter RNA transcripts are polyadenylated (**Fig 14b**). No RT controls, indicated that these products were from RNA and not contaminating genomic DNA. For these experiments, I used DNase I from another vendor (Worthington) which proved to be far more effective and could reduce genomic DNA below detectable levels with one treatment and leaving no need for restriction enzyme treatment. Purified genomic DNA was used for a positive control (**Fig 14b**).

In order to measure the relative abundance of the promoter RNA transcript with respect to the PR mRNA, I used quantitative real time PCR (qRT-PCR). I designed primers to the PR promoter and across every exon boundary (**Fig 14a**). All primers were tested for linear amplification by standard curve where serial dilutions of cDNA are amplified and plotted as C_T versus $\log_{10}$ [cDNA] (*279, 280*). Linear amplification is defined if the slope of the plot lies between -3.0 and -3.6.

qRT-PCR of polyA purified RNA from T47D or MCF7 cells revealed relatively even levels of cDNA across the PR mRNA (n = 6 for each cell line). The promoter RNA ranged from 10 to 1000 fold lower than the PR mRNA (**Fig 14c-d**). Although this result was confusing at the time, it later proved to be indicative of more than one RNA species existing in the PR promoter but at different relative abundances.

In order to discover the sequence of the RNA transcripts in the PR promoter I used rapid amplification of cDNA ends (RACE). RACE can be modified to clone 5' ends or 3' ends. I used a GeneRacer® kit (Invitrogen). RACE ligates a known sequence (linker) to either the 5' or 3' ends and uses reverse transcriptase PCR to amplify the cDNA end with one primer targeting the linker and one primer within the region of interest. For this protocol, RNA was treated with a phophatase before decapping to ensure that degraded RNA fragments are not cloned. 3' RACE cDNA is synthesized using oligodT conjugated linker to ensure only polyadenylated RNAs are cloned.

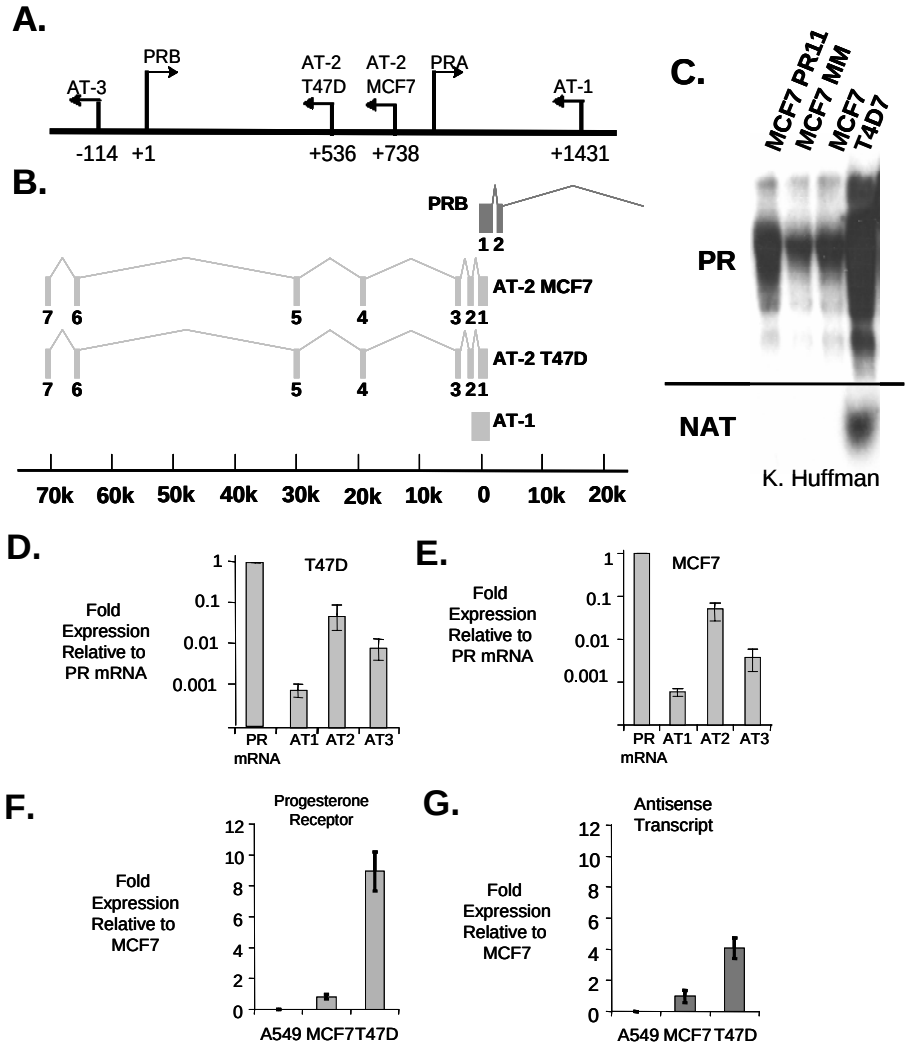


Figure 16: Indentification of antisense transcripts overlapping the PR promoter. (A) Four unique transcription start sites were identified by 5' RACE. (B) Start sites AT-1 and AT-2 produced transcripts overlapping the region targeted by agRNAs, one spliced and one unspliced. (C) The transcript AT-2 could be detected by Northern blotting in T47D cells but was too low abundance to be detected in MCF7 cells. Knowing the sequence of each transcript allowed specific primers to be used for qRT-PCR to measure the relative levels of the antisense transcripts in T47D (D) or MCF7 (E) cells. Both the PR mRNA (F) and antisense transcript AT-2 (G) is higher expressed in T47D cells than MCF7. Neither is detected in lung A549 cells. Error bars are standard deviation (n = 6).

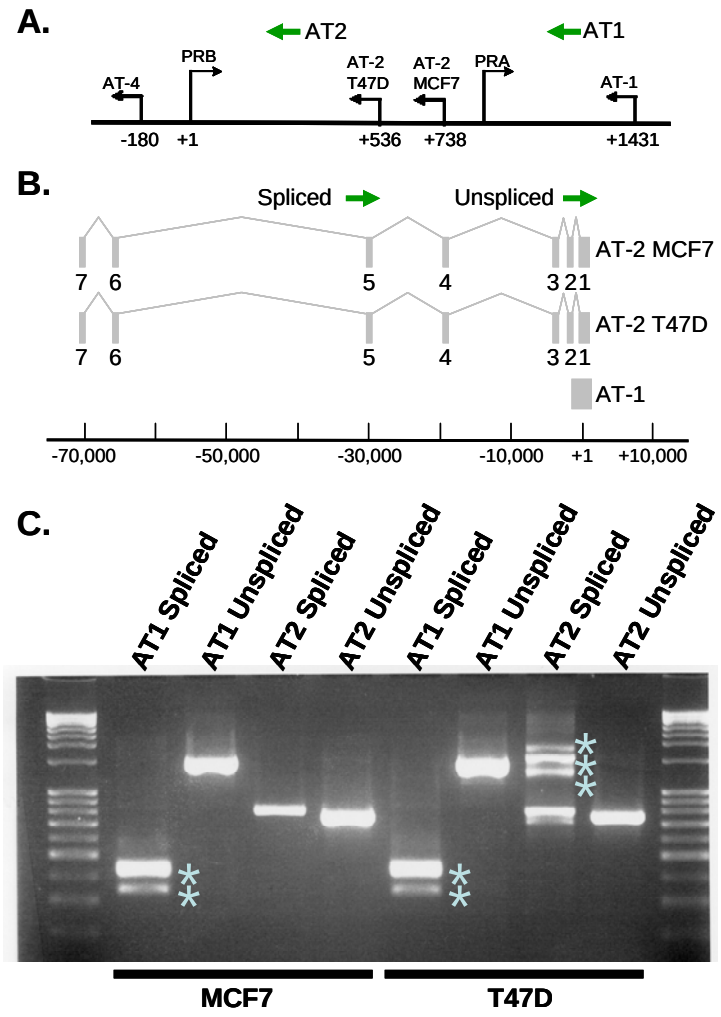


Figure 17: Uniting 5' and 3' RACE products. (A) Primers were designed to distinguish the more proximal (AT2) and more distal (AT1) antisense transcription start sites (TSS) with respect to PRB TSS. (B) Reverse primers were designed to distinguish spliced or unspliced transcripts. (B) No product was detected for a spliced transcript originating from the AT1 TSS. Product was detected and sequenced for an unspliced product transcribed from AT1 and AT2 TSSs which could be explained by a single unspliced product originating from the AT1 TSS. Product was detected and sequenced for a spliced transcript originating from AT2. (*) Indicates nonspecific PCR product verified by cloning and sequencing.

A.

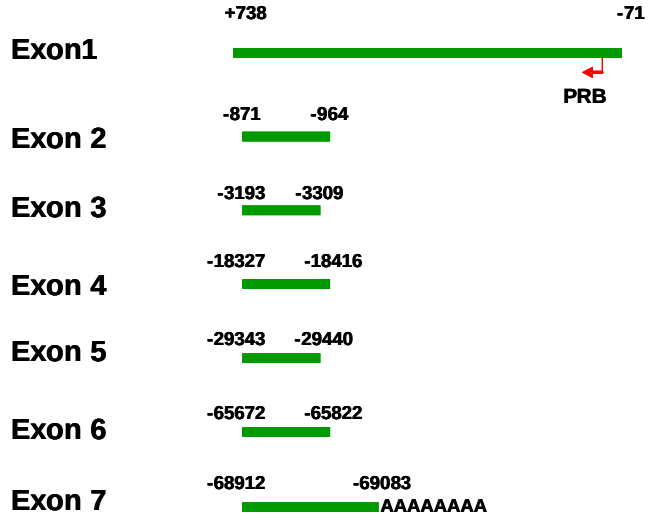
AT1

GGACTCAGAGCCATCCTCCTCCTCAACCTCCACCGCAGCGGCCTGCGGAGACGGCTTCACTG
GGGCCCCGGACCAGTGAGGGCTCTCAGAGGCCGGGAGCAGCAGCTGCCGGGCTGGTGACAG
GCCCCGGGGCAGCACTTTATGGGCAGCTGCCGTCCCGGAGCTGTCTCCAACCTTGCACCCGG
ACCGGCTCATGAGCGGGGACAACACCCGCTGGGTGGCGGGGGCAGCCGGTGGATCTTCGGG
AAGTTCGGGGCCAAACAGGCACCAAGAGCTGGTGACCTCGCAGGCGGGAGGGCTGGGTTGG
CTCTGCCCGGGACCTGAGGGCGCCAACAGAGTGTCCAAGACACTGTCCAGCAGTCCGCTGTC
CTTTTCTGGGGGACTAGAACTGCTGCCTCCAGCACCCCTTGTAGCTTCAGCTCTGGAATATGC
GCCCTCCACGTCCGACAGCGACTGCTGGTCCTGCGTCTTTTCGTCGGAGGGGTCCTGTCCCTG
GCAGGGCCGAGGGAAGAGTAGCCCGTCCAGGGAGATAGGTATGGCCGAAACTTCAGGCAAG
GTGTCCGAGGTCTGGCTCCCCGGGAACGGACCTGCGGCTGGGCGACACAGCAGTGGGGATCC
GACCTCGGGGGAGGGCGGGCCGCCCGCCACGTGGGGAGCCCGGGGACCCTTTGCCTTCAGCT
CAGTCATGACGACTGGACTCCCCTTTTCTCCTCCCCCGTCTCCAGGAGGAGGGAAAAGGGAA
GGAGGAGGGGGTTTCGGGAATATAGGGGCAGAGGGAGGAGAAAGTGGGTGTTGAATGTGGC
TGGACCGGAGGGATCTCCACCTCCTGGGTCGGGGGCGGGGGAGGGCGGCGCTGGTCAGCTC
CTGCCCTTGGCCTCCATCCTGTCGTCAGGGGAACTGTGGCTGTCGTTTGTCCCAGCGAGCGGC
AAGTGGGGAGCGCAAGAAAAAGTAGTAATTGTTAGGAGATCTCGTCTCCTAACTCGGGGAGT
TCTCCAAGAGAGTTCTCCAACTTCTGTCCGAGGACTGGAGACGCAGAGTACTCACAAGTCCG
GCACTTGAGTGGCTGCGGCTGCGACGGCAATTTAGTGACACGCGGCTCCTTTATCTCCCGACT
TTTTCTCTGGCATCAAACTCGTGCATGCTGTGAAGCTCTCAGTCCCTCGCTGAGTTCCACTGC
CCCCTCACTAAAACCCTGGGGCTAGTCGGACCTCTCGGTACAGCCCATTCCCAGGAAGGGTC
GGACTTCTGCTGGCTCCGTACTGCGGGCGACAGTCATCTCCGAAGATCTCAGATCCCAGTAG
TGCGGGAGCACTAGCCGCCTCGGGTTGTAGATTTCACTCAAATGACAAGTGAAGCTAGTTCT
CATTGAGAATGCCACCCACACGCACAAATACAACAAGGCTTACCCCGATTAGNGACAGNTGT
GGAC**TNNC**CAGACAGNTTTNTAACAANGCCTCCTCNTCTAGGGNNGNCCCGCCCAAAGCCC
CTCCCTACCCCAATTACCNNNAGGATCTGAAACTCTGGAGTTGGCATTTCCACCCGTTATTCT
GAATGCTACTCTCAATAGCAGGTTCTTTGGGATGGAACCTCATAAGCATATTACGTTTTGTTT
TGCAAATTAAGAATTATGCCCTATCTAATTGNAAAAATGAATAGATTCTATCAGAAGTAGAA
TTTTTGTCACCATTTTAAGATTTCAGTTTTGTAAAGATTTAACACAGAGGAAGACATTTGGCT
ACATTATCTTTTAAAATAATAAATGTATCAAGGACGATACTAAATAAGAGATTCTCTCCCTTA
TGAGTTCCATAAAAGTGAATGCTTTCAAGTTTCTCCTGCTGCGACCAACATGTCCTGCTCCTC
TTGAGTATCCTCAGCGTGACAAAGAAATTTGNGAGTGGGAGATGGTAACGCCTCTGTATCTA
TGATGGAAGTTTGNATGTTGTGTGCCACACTTCGATTTGTCTTAAGGAATGNGTTCCAATTTT
TAGTAAATAGCACTTTAANGNAAGTTTCCTTTAGCTCCAGCTTATACAGTGNGCTNCATCATA
ACATCCTCAGNAGAGATGTTGAACAGTACTTTTTNCCCTANATATTCTTAGNCTGAAANGTA
ANTATCNTTGAGCTTATAAAATGGNTCCCCGCTAAATTCATGAANNCANNAAAAAAAAAAA
AAAAAAAA

B.

AT2-MCF7

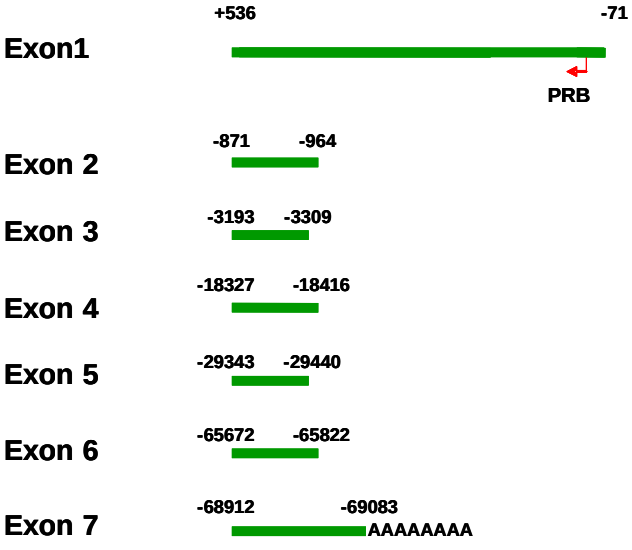
TGGACTCCCCTTTTCTCCTCCCCCGTCTCCAGGAGGAGGGAAAAGGGAAGGAGGAGGGGGTT
TCGGGAATATAGGGGCAGAGGGAGGAGAAAGTGGGTGTTGAATGTGGCTGGACCGGAGGGA
TCTCCACCTCCTGGGTCGGGGGCGGGGGAGGGCGGCGCTGGTCAGCTCCTGCCCTTGGCCTC
CATCCTGTCGTCAGGGGAACTGTGGCTGTCGTTTGTCCCAGCGAGCGGCAAGTGGGGAGCGC
AAGAAAAAGTAGTAATTGTTAGGAGATCTCGTCTCCTAACTCGGGGAGTTCTCCAAGAGAGT
TCTCCAACTTCTGTCCGAGGACTGGAGACGCAGAGTACTCACAAGTCCGGCACTTGAGTGGC
TGCGGCTGCGACGGCAATTTAGTGACACGCGGCTCCTTTATCTCCCGACTTTTTCTCTGGCAT
CAAACTCGTGCATGCTGTGAAGCTCTCAGTCCCTCGCTGAGTTCCACTGCCCCCTCACTAAAA
CCCTGGGGCTAGTCGGACCTCTCGGTACAGCCCATTCCCAGGAAGGGTCGGACTTCTGCTGG
CTCCGTACTGCGGGCGACAGTCATCTCCGAAGATCTCAGATCCCAGTAGTGCGGGAGCACTA
GCCGCCTCGGGTTGTAGATTTCACTCAAATGACAAGTGAAGCTAGTTCTCATTGAGAATGCC
ACCCACACGCACAAATACAACAAGGCTTACCCCGATTAGNGACAGNTGTGGAC**TNNC**CAGA
CAGNTTTNTAACAANGCCTCCTCNTCTAGGGNNGNCCCGCCCAAAGCCCCTCCCTACCCCAA
TTACCGGACCTCAAGGTCTAGCTGTGCTAATGACTCATAGTTTATNTCANCCATGTATAAAGA
ATGCAGAAGACTCCAGAAGGTGGGGGAGCCACTAGAGGATTCCATCCAGGACACCACATTT
AATTGTTATGATGTCTTGTGCTCCTCTTGGCTGTGAGAGTTTCTCAGATTTTCCTTGTATTTGA
TGACCTTGACAGTCCTTAGGAGTACTGAGATGGGGGTTTCACGATATTGCCCAGGCTGGTCCC
AAACTCCTGGCCTCAAGCTATCCTCCTGCCTTGACGTTCGAAAGCACTNAGATTTTCCTNTCT
GTCCTGCCGCCATGTGAAGAAANATGTGTTTGCTTCCCCTTCCNCCGTGATTGTANTTTTCCT
GCAGCCTCCCCAGCCAAGCTGAACTGCAATGAGGAAGCAATAAAGGACTNTGAGCAAGAGA
ACGATGTATATAGGCCATTGTTTTGGAAGATTCATCNCAAATTCATGTGCAAGTAGATTGGA
GGAAGGAAGTATCAACANAAANAAAGGCCAATTATGAGACCATTGCAATACTATATGATGA
GAGCATCAGAGNCAATAACCAAGGTNTTATAGATACCAAGTGAGGAATCTGGGATTANAAG
TCAGTNTATTTCATTGGAAAGGCTGTTTTCAGTNTTTTTCACTGGAAAAGTTGTCATCCTGTGT
CTTTNTTATAGTACATATATTNCAGTAATAAAACTTTATTTTCCTTTTCAAAAAAAAAAAAAA
AAAAAA



C.

AT2-T47D

AACTGTGGCTGTCGTTTGTCCCAGCGAGCGGCAAGTGGGGAGCGCAAGAAAAAGTAGTAATT
GTTAGGAGATCTCGTCTCCTAACTCGGGGAGTTCTCCAAGAGAGTTCTCCAACTTCTGTCCGA
GGACTGGAGACGCAGAGTACTCACAAGTCCGGCACTTGAGTGGCTGCGGCTGCGACGGCAAT
TTAGTGACACGCGGCTCCTTTATCTCCCGACTTTTTCTCTGGCATCAAACTCGTGCATGCTGTG
AAGCTCTCAGTCCCTCGCTGAGTTCCACTGCCCCCTCACTAAAACCCTGGGGCTAGTCGGACC
TCTCGGTACAGCCCATTCCCAGGAAGGGTCGGACTTCTGCTGGCTCCGTACTGCGGGCGACA
GTCATCTCCGAAGATCTCAGATCCCAGTAGTGCGGGAGCACTAGCCGCCTCGGGTTGTAGAT
TTCACTCAAATGACAAGTGAAGCTAGTTCTCATTGAGAATGCCACCCACACGCACAAATACA
ACAAGGCTTACCCCGATTAGNGACAGNTGTGGAC**TNNC**CAGACAGNTTTNTAACAANGCCT
CCTCNTCTAGGGNNGNCCCGCCCAAAGCCCCTCCCTACCCCAATTACCGGACCTCAAGGTCT
AGCTGTGCTAATGACTCATAGTTTATNTCANCCATGTATAAAGAATGCAGAAGACTCCAGAA
GGTGGGGGAGCCACTAGAGGATTCCATCCAGGACACCACATTTAATTGTTATGATGTCTTGT
GCTCCTCTTGGCTGTGAGAGTTTCTCAGATTTTCCTTGTATTTGATGACCTTGACAGTCCTTAG
GAGTACTGAGATGGGGGTTTCACGATATTGCCCAGGCTGGTCCCAAACTCCTGGCCTCAAGC
TATCCTCCTGCCTTGACGTTCGAAAGCACTNAGATTTTCCTNTCTGTCCTGCCGCCATGTGAA
GAAANATGTGTTTGCTTCCCCTTCCNCCGTGATTGTANTTTTCCTGCAGCCTCCCCAGCCAAG
CTGAACTGCAATGAGGAAGCAATAAAGGACTNTGAGCAAGAGAACGATGTATATAGGCCAT
TGTTTTGGAAGATTCATCNCAAATTCATGTGCAAGTAGATTGGAGGAAGGAAGTATCAACAN
AAANAAAGGCCAATTATGAGACCATTGCAATACTATATGATGAGAGCATCAGAGNCAATAA
CCAAGGTNTTATAGATACCAAGTGAGGAATCTGGGATTANAAGTCAGTNTATTTCATTGGAA
AGGCTGTTTTCAGTNTTTTTCACTGGAAAAGTTGTCATCCTGTGTCTTTNTTATAGTACATATA
TTNCAGTAATAAAACTTTATTTTCCTTTTCAAAAAAAAAAAAAAAAAAAA



D.

AT-3 5' End

TTCTGAATGCTACTCTCAATAGCAGGTTCTTTGGGATGGAACCTCATAAG
CATATTACGTTTTGTTTTGCAAATTAAGAATTATGCCCTATCTAATTGGA
AAAATGAATAGATTCTATCAGAAGTAGAATTTTTGTCACCATTTTAAGAT
TTCAGTTTTGTAAAGATTTAACACAGAGGAAGACATTTGGCTACATTATC
TTTTAAAATAATAAATGTATCAAGGACGATACTAAATAAGAGATTCTCTC
CCTTATGAGTTCCATAAAAGTGAATGCTTTCAAGTTTCTCCTGCTGCGAC
CAACATGTCCTGCTCCTCTTGAGTATCCTCAGCGTGACAAAGAAATTTGG
GAGTGGGAGATGGTAACACCTCTGTATCTATGATGGAAGTTTGGATGTTG...

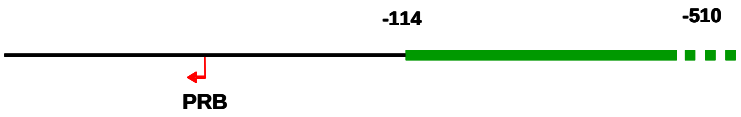


Figure 18: Sequence of four transcripts overlapping the PR promoter. (A) AT-1 is an unspliced transcript overlapping the PR promoter. (B) AT-2 MCF7 is a spliced transcript overlapping the PR promoter in MCF7 cells. (C) AT-2 T47D is a spliced transcript overlapping the PR promoter in T47D cells. (D) AT-3 is only the 5' end of a transcript that only partially overlaps the PR promoter.

To detect either 3' ends of sense transcripts or 5' ends of antisense transcripts, I used every possible combination of the forward primers targeting sites -628, -520, -150, -77, -33, +33, and +102 with respect to the PR transcription start site (**Fig 14a**). To detect either 5' ends of sense transcripts or 3' ends of antisense transcripts, I used every possible combination of the reverse primers targeting sites +250, +199, +113, -10, and -388. 5' and 3' RACE was performed using cDNA samples prepared from either MCF7 or T47D cells (**Fig 14a**).

Using this protocol, 5' and 3' RACE recovered four unique antisense transcription start sites and two unique 3' ends (**Fig 15b, 16a-b**). Transcript AT-1 overlapped the PR mRNA the most by 1400 nucleotides. Transcript AT-2 had different transcript start sites in T47D and MCF7 cells, separated by 200 nucleotides and overlapped the PR mRNA by either 530 or 730 nucleotides respectively. Transcript AT-3's 5' end began 113 nucleotides upstream of the PR transcription start site and did not overlap the region targeted by agRNAs so I did not pursue identifying its 3' end.

To verify which transcription start site corresponded to which 3' end I used long RT-PCR (*204*). I designed forward primers to target near the transcription start site for AT-1 and for AT-2 T47D (**Fig 17a**). I placed reverse primers to detect either the unspliced transcript at -432 nucleotides upstream of the PR transcription start site or in exon 5 of the spliced transcript (**Fig 17b**). Primers directed at transcript AT-1 produced PCR product of the correct molecular weight for an unspliced transcript but not for a spliced transcript. Primers directed at the transcript AT-2 produced PCR product of the correct molecular weight for a spliced transcript and also for an unspliced transcript since these primers would also detect an unspliced transcript originating for AT-1 (**Fig 17c)**. In both MCF7 and T47D, a low molecular weight band is seen for the primer detecting AT-1 and the exon 5 primer. These bands were cloned, sequenced, and found to be non-specific PCR product. In T47D, additional higher molecular weight bands were seen for the spliced AT-2 transcript. These bands also were cloned and sequenced and found to be nonspecific PCR product (**Fig 17c**).

Knowing the exact sequence, primers could be used to distinguish the relative abundance of each RNA transcript by qRT-PCR revealing the spliced transcript AT-2 to

be the most highly abundant (10 to 20 fold lower than PR mRNA), followed by the transcript AT-3 (100 fold lower than PR mRNA), and the unspliced transcript AT-1 was to lowest expressed (more than 1000 fold lower than PR mRNA) (**Fig 16d-e**). Concentrations of the antisense transcript rose proportionally to PR mRNA for T47D and MCF7 and neither was detectable in lung cancer A549 cells (**Fig 16f-g**).

To validate that the transcript AT-2 exists, I collaborated with K. Huffman to generate a Northern analysis. I purified RNA from T47D cells, MCF7 cells, and MCF7 cells treated with activating agRNA PR11. 5 μg of total RNA was run on a denaturing gel, and blots incubated with a probe targeting exon 7/8 of the PR mRNA overnight. Blots were exposed to film for 3 days to show PR mRNA, which was higher in T47D and in PR11 treated MCF7 cells than in untreated MCF7 cells (**Fig 16c**). The membrane was stripped and hybridized with a probe targeting exon boundary 2/3 of the transcript AT-2. This was developed for 3 days and a band was detected for T47D cells but not for MCF7 cells, presumably because the transcript is too lowly expressed (**Fig 16c**).

4.3 agRNA ACTIVATION CHANGES THE TRANSCRIPTIONAL LANDSCAPE

Previously the Corey lab had contracted the Center for Functional Genomics at University of Albany to perform 5' RACE for PR to validate the transcription start site for PR-B. For this experiment, 20 clones were sequenced to reveal somewhat of a distribution of transcription start sites. This observation raised the possibility that some rare upstream transcription start site might produce a target for agRNAs but that it was overlooked because of its rarity with respect to the major transcription start site. To address this question, I performed 5' RACE on the PR-B mRNA and sequenced 60 clones for each cell line, T47D and MCF7, using reverse primers at +250 and +199.

I used a RACE protocol that selects for capped RNAs to ensure degraded RNAs are not sequenced. A single transcription start site is not seen. In T47D, 37% (22 out of 60 clones) of transcription start sites lie on or within 1 nucleotide of the previous published major transcription start sites for PR-B at +1 and +15 (**Fig 19a**) (*281*). The rest of the transcription start sites lie largely clustered between +75 and +110. For MCF7,

64% (40 out of 62 clones) lie at +1 and +15 with fewer transcripts beginning downstream (**Fig 19b**). I also performed 5' RACE on MCF7 cells treated with the activating RNA PR-11. For these cells, a shift in the distribution of transcription start sites is seen with only 26% (14 out of 54 clones) of transcription start sites appearing at +1 one and +15 and the majority of sites appearing downstream between +75 and +110 (**Fig 19c**). Even by gel electrophoresis on a 3% agarose gel, two bands were distinguishable (**Fig 19d-e**). Cutting each band separately revealed that the upper band is enriched for the +1 and +15 start sites and the lower band is enriched for the downstream transcription start sites. This observation agrees with the observation of a distribution around two clusters of transcription start sites.

Out of a total of 176 clones sequenced from MCF7 and T47D cells, not one transcription start site is seen upstream of the previously reported transcription start site. Even after placing a primer bracketing the transcription start site to bias the PCR to detect upstream transcription start sites, no sites were detected (**Fig 19d-e**). This strongly suggests that no upstream variant of the PR mRNA can serve as a target for agRNAs.

Since 5' RACE begins with RNA treated with the phosphatase, CIP, followed by the decapping enzyme,TAP, the downstream transcription start sites seen by RACE are from capped transcripts and not degraded RNAs. The fact that activating MCF7 cells with PR-11 drives the distribution of transcription start sites more like T47D's distribution, which also highly expresses PR, is intriguing and possibly hints at a biological significance behind the appearance of the alternative RNA transcripts.

On the other hand, Kastner et al. does not report detecting any downstream transcription start sites by S1 nuclease mapping or by RNase protection (*281*). Kastner et al. uses a S1 probe extending to +191 so the downstream start sites should have been visible. Kastner used polyA purified RNA suggesting that perhaps the downstream transcription start sites are not full length polyadenylated transcripts. But Kastner did not show the gel for these sites, so faint bands may have been present. Both PCR and cloning are more efficient at detecting shorter transcripts. So it is likely that the downstream transcription start sites are rarer than they appear by the distributions in **Fig 19**. However, the shift in distribution is hard to dismiss.

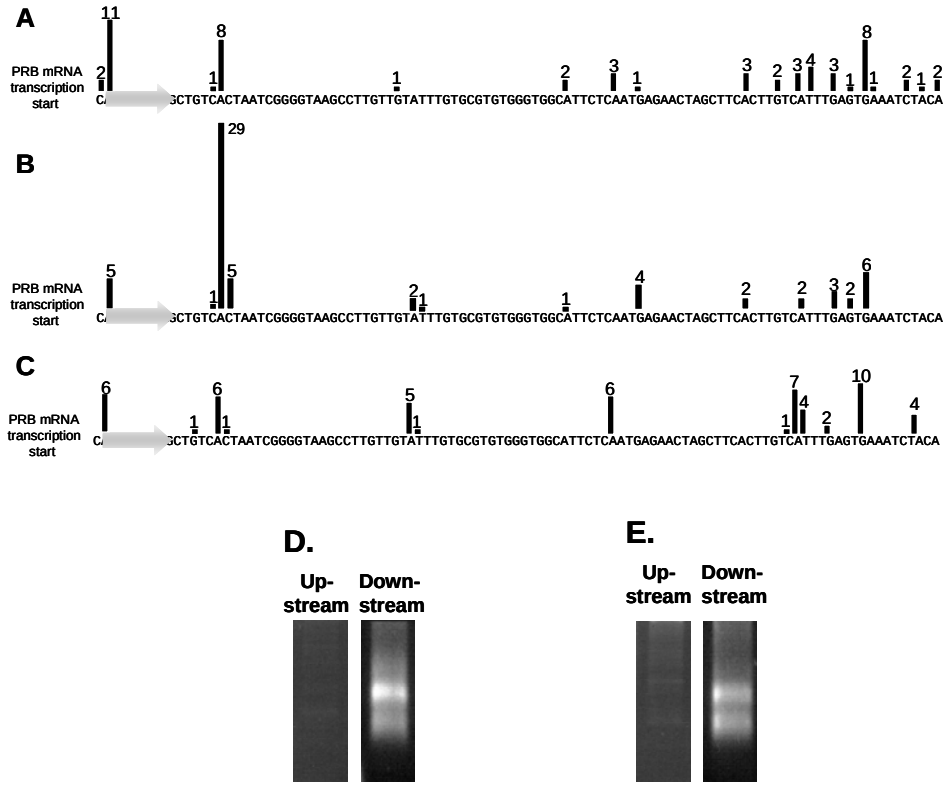


Figure 19: The distribution of transcription start sites for PR-B. (A) Several transcription start sites (TSS) are seen for PR-B in T47D cells (n=60). Above each nucleotide is written the number of clones sequenced with that TSS. However, no TSSs are seen upstream of the annotated transcription start site for PR-B. (B) A different distribution of TSSs is seen for PR-B in MCF7 cells, which expresses constitutively lower levels of PR (n=62). (C) When treated with activating agRNA PR-11, the distribution of TSSs changes for PR-B to be more like the distribution for T47D (n=54). (D) 5' RACE product was ran on a 1.5% gel. Primers targeting downstream of the TSS detects the T47D TSS, but an upstream primer bracketing the TSS detects no upstream TSSs for PR-B. (E) Similarly, primers targeting downstream of the TSS detects the MCF7 TSS, but an upstream primer bracketing the TSS detects no upstream TSSs for PR-B.

There have been several reports of bimodal transcription around transcription start sites, but my 5' RACE experiments using primers near the transcription start site designed to detect antisense transcripts (at -150, -77, -33, and +33) did not amplify short antisense transcripts (*205, 214-217, 282*). However, published reports of short transcripts suggest them to be quite rare so it is possible that they may be too rare for 5' RACE to detect. Also reported are short RNAs originating from downstream of transcription start sites (*216, 282*). These small RNAs are suggested to be degraded products from mRNAs but somehow are capped with a protective 5' structure sensitive to treatment with TAP (*216*). This is another possibility to explain these downstream transcription start sites that they are not full length transcripts but short degraded products of mRNAs that accumulate when PR is highly expressed. Performing 5' RACE with primers further downstream or with polyA purified RNA might differentiate these possibilities.

4.4 THE NONCODING RNA IS NECESSARY FOR agRNA ACTIVATION

To test if the antisense transcript is necessary for agRNA activity, I designed antisense oligonucleotides called "gapmers" to knockdown of the antisense transcript. Gapmers are a single stranded mixed synthetic oligonucleotide with a DNA core and 2' MOE flanks to provide nuclease resistance. Gapmers bind their RNA targets and recruit RNase H to destroy the RNA target (**Fig 20**). I chose to use 2'-MOE modified oligonucleotides because R. Ram had shown that they were not effective at targeting genomic DNA and silencing transcription, which would confound the interpretation of gapmer results (*236*).

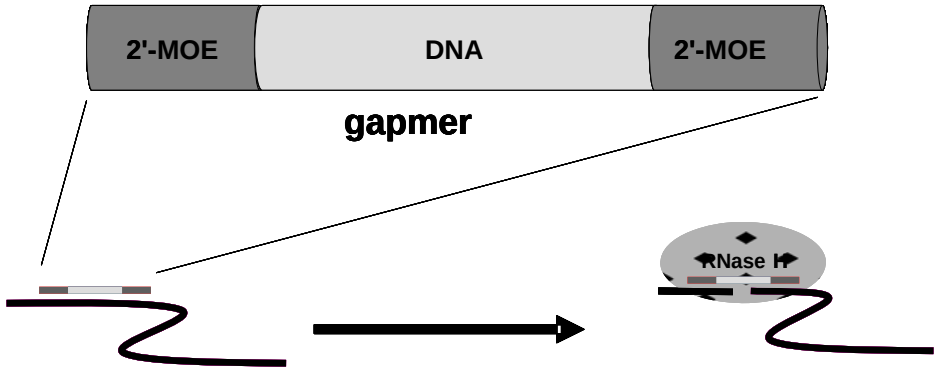
Gapmers were synthesized by B. Monia at ISIS pharmaceutics. Ten gapmers were designed with 5 targeting a putative sense transcript and 5 targeting a putative antisense transcript (**Fig 20**). At the time the gapmers were designed, I did not know the sequence of the antisense transcripts. Two more gapmers were designed to target the PR-B and PR-A mRNA. B. Nguyen transfected gapmers into MCF7 cells in 6-well dishes with a concentration of 50 nM using RNAi Max lipofectamine and harvested 5 days later (*283*). At 50 nM gapmers were toxic to cells, seen by a dramatic slowing of growth.

Despite the toxicity, significant changes to PR expression were not seen by western (**Fig 21c**). Unfortunately, the positive control gapmers targeting the PR-B mRNA and the PR-A mRNA did not knockdown PR expression.

Gapmer concentrations were reduced to 25 nM. B. Nguyen and I worked together to transfect and harvest MCF7 cells. We measured the knockdown by qRT-PCR for all of the gapmers except for the gapmers targeting -102 which were very toxic. Several gapmers induced nonspecific knockdown of the antisense transcript even though they did not target the antisense transcript (**Fig 21a**). To measure knockdown I used qRT-PCR primers spanning -33 to +199, which could detect either AT-1 or AT-2 or even any other RNA transcript that I may not have discovered by 5' RACE. Gapmers did not change PR expression seen by qRT-PCR of the mRNA in MCF7 cells (**Fig 21b**). I decided to continue experiments with two antisense targeting (-20 and -32) and one sense targeting gapmer which I renamed gapmers 1, 2, and 3.

B. Nyugen and I transfected and harvested MCF7 and T47D cells. We developed a transfection protocol so that cells were transfected with 25 nM gapmers or agRNA on day 1 using RNAi-Max (Invitrogen), then MCF7 cells were split 1:3 and T47D cells split 1:2 on day 4 and reseeded into 6-well dishes. Upon reseeding, cells were reverse transfected with 25 nM gapmer and 25 nM agRNA with RNAi-Max (Invitrogen). Cells were harvested on day 7 for RNA using Trizol® reagent (Invitrogen). Reverse transfection was performed basically as described by R. Beane.

Again, for a single 25 nM transfection, gapmer G1 reduced antisense transcript levels, G2 reduced antisense transcript levels to a lesser degree or not at all, and G3 had no effect on antisense transcript expression (**Fig 22a-b**). Again, no change were seen to PR expression in either MCF7 or T47D after gapmer knockdown of the antisense transcript (**Fig 22c-d**). Using the double transfection method, efficient knockdown of the antisense transcript and reversal of gene activation was seen in MCF7 cells for gapmer G1 (**Fig 22e**). For T47D cells, reversal of agRNA silencing was not achieved (**Fig 22f**). This was due to the fact that gapmers did not effectively silence the antisense transcript in the double transfection protocol (**Fig 22h**).



ISIS pharmaceuticals, Carlsbad CA

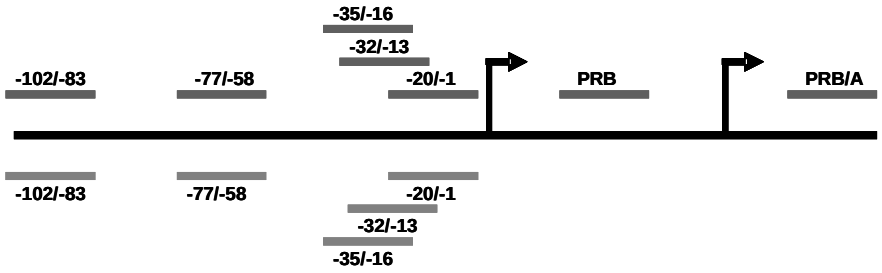


Figure 20: Targeting noncoding RNAs with antisense gapmers. (Top) Gapmers are designed to have a DNA central section with 2'-MOE flanks to enhance nuclease resistance. Gapmers bind their RNA targets forming an RNA:DNA hybrid that is recognized by RNase H, which cleaves the RNA transcript. (Bottom) 12 gapmers were designed. 5 targeting an antisense transcript, 5 targeting a sense transcript, and 2 positive controls targeting the PR-B and PR-A mRNA.

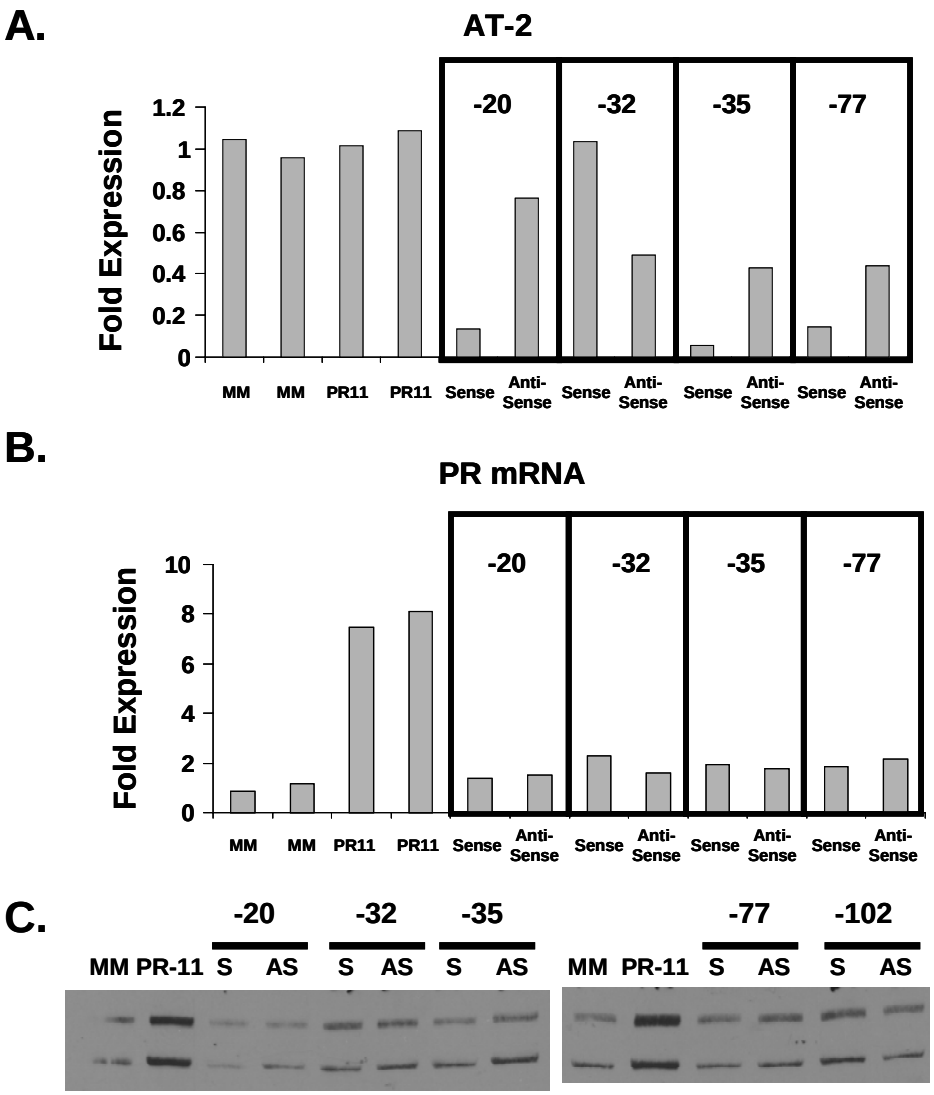


Figure 21: Gapmers knockdown the promoter RNA but do not effect PR expression. (A) Gapmers were screened at 50 nM in MCF7 cells. qRT-PCR of transcript AT-2 revealed several gapmers knocked down the antisense transcript. Several sense targeting gapmers nonspecifically knocked down the antisense transcript. (B) No gapmers had an effect on PR mRNA. (C) Western blot reveals that no gapmer had any effect on PR protein levels.

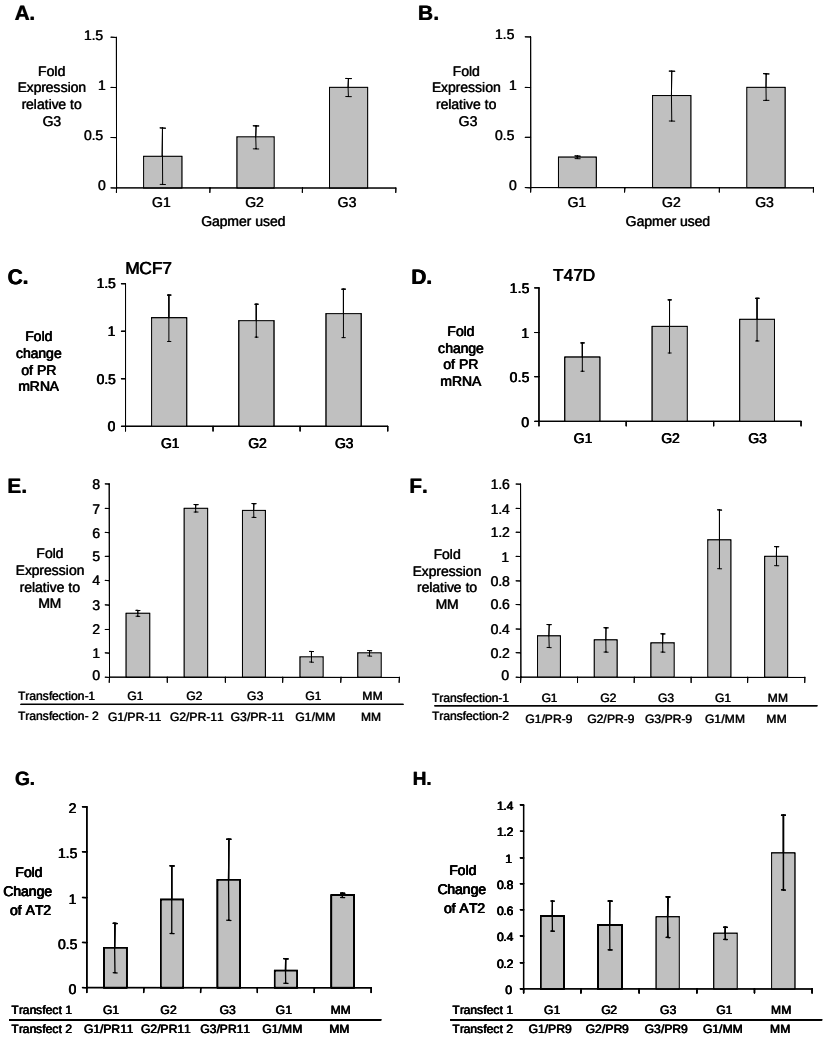


Figure 22: Gapmers reverse gene activation. Using 25 nM transfections, qRT-PCR reveals that gapmers G1 and G2 in MCF7 (A) and G1 in T47D (B) could knockdown the antisense transcript. Gapmer knockdown of the antisense transcript had no effect on PR expression in MCF7 (C) or T47D (D). Gapmer knockdown of the antisense transcript could reverse gene activation in MCF7 (E) but was unable to effect silencing in T47D cells (F). Knockdown of the antisense transcript was strong in MCF7 (G) but weak in T47D (H) in the double transfection strategy. This may why reversal was not accomplished in T47D. Data is averaged from 3 independent experiments and error bars represent standard deviation.

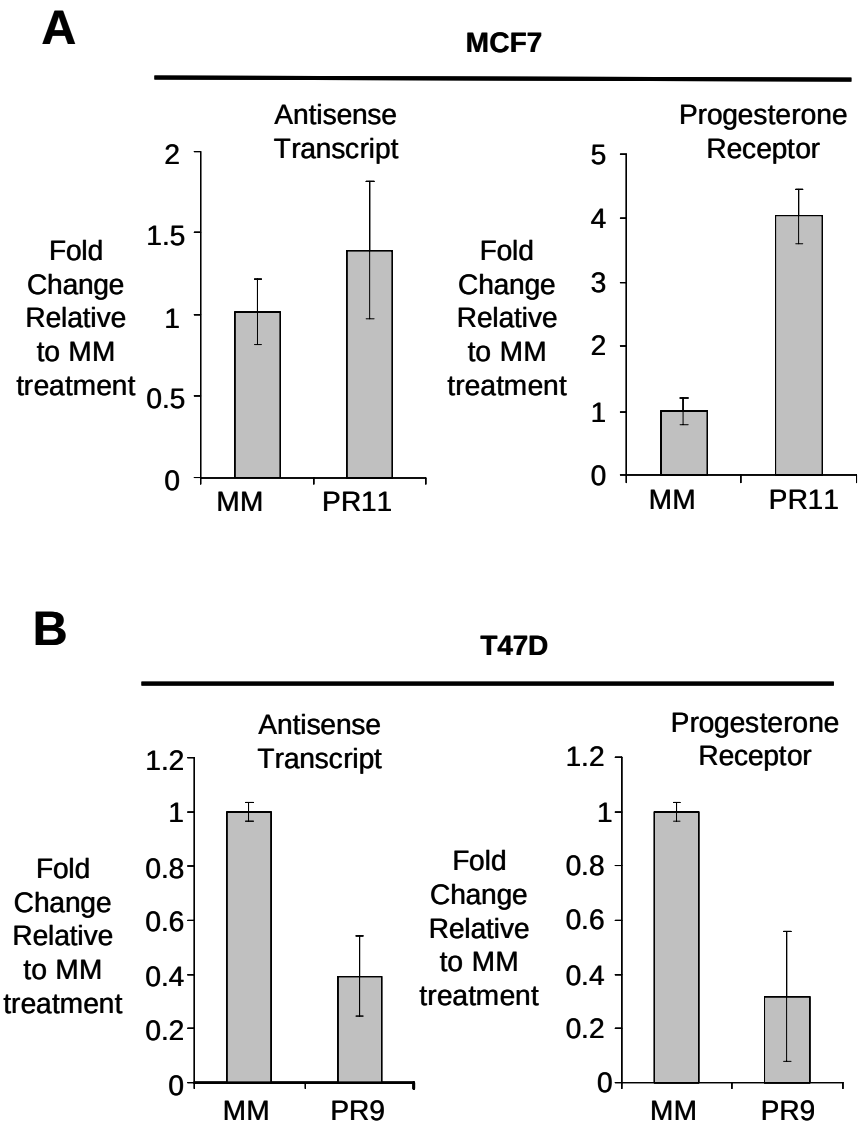


Figure 23: The antisense transcript levels under treatment with agRNAs. (A) In MCF7 cells, qRT-PCR reveals that treatment with activating agRNA PR-11 has no effect on the expression levels of the antisense transcript. (B) In T47D cells, silencing agRNA PR-9 also silences the antisense transcript. This silencing may be at the level of transcription and not due to RNA cleavage.

There are many reasons that gapmers could not reverse agRNA silencing. Gapmers must, without the aide of proteins to guide them to their target, compete with agRNAs, which do have proteins to help them find and hybridize their targets. The fact that gapmer knockdown was not more than typical agRNA knockdown of the antisense transcript in double transfected T47D cells suggests that gapmers in this double transfection strategy did not work effectively at all. However, reversal by gapmers is one of three strategies used to implicate antisense transcripts in the agRNA mechanism in this study.

The reversal of agRNA activation by loss of the antisense transcript suggests that the antisense transcript is necessary for agRNA activity. This result agrees well with model 2 that the antisense transcript serves as a substrate for agRNA recognition of gene promoters (**Fig 12**). Recently, Morris et al. have suggested that gene activation is achieved to cleavage of antisense transcripts that negatively regulate gene expression (*284*). However, this conclusion is contradicted by the fact that cleavage of the antisense transcript by gapmers had no effect on PR expression in either MCF7 or T47D cells (**Fig 21, Fig 22c-d**). This argues that the antisense transcript has no intrinsic regulatory role in PR expression for MCF7 and T47D cells. Furthermore, treatment of MCF7 cells with PR-11 has no effect on antisense transcript levels (**Fig 23a**), ruling out cleavage of the transcript as part of the mechanism. Treatment of T47D cells with PR-9 does silence the antisense transcript to some extent but this silencing may be at the level of transcription and not due to cleavage (**Fig 23b**).

4.5 agRNAS ASSOCIATE WITH THE ANTISENSE TRANSCRIPT

To test if agRNAs physically interact with genomic DNA or with the antisense RNA transcript, S. Younger performed an experiment using 3'-biotinylated agRNAs (Sigma-Proligo). RNA duplexes with a 3' biotin on either the sense or antisense strand were transfected at 100 nM into T47D and MCF7 cells grown in six-well dishes. Biotinylated PR-11 activated as well as unmodified in MCF7 cells (**Fig 24a-b**).

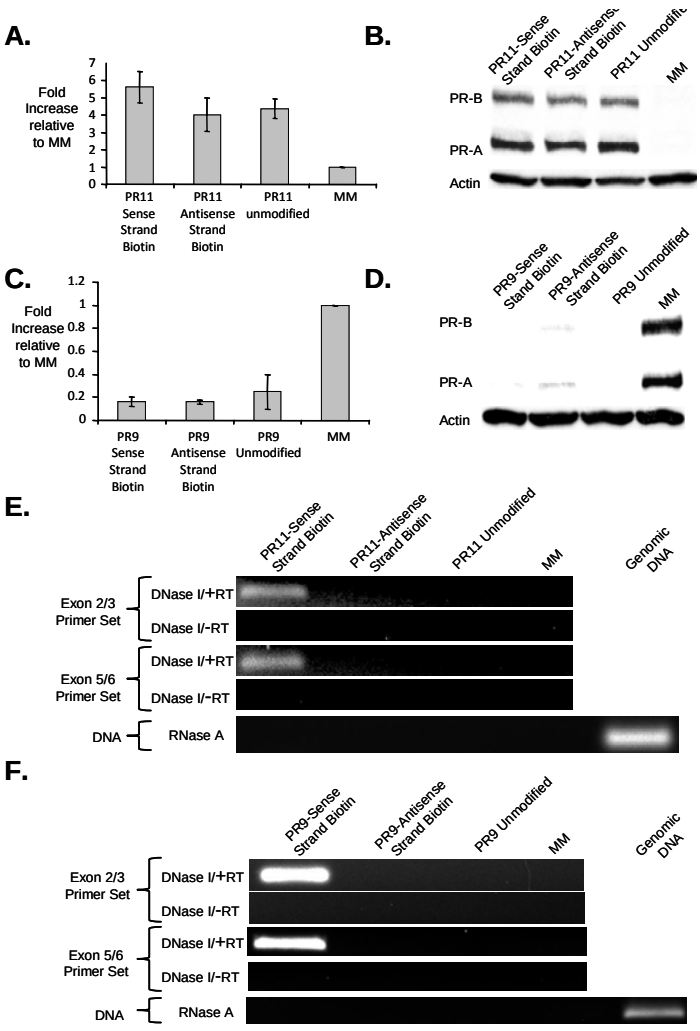
Biotinylated PR-9 silenced as well as unmodified in T47D cells (**Fig 24c-d**). This was seen at the protein level, by western, and at the mRNA level, by qRT-PCR.

Treatment of recovered RNA with reverse transcriptase allowed the amplification of the antisense transcript AT-2 by primers targeting either exon boundary 2/3 or exon boundary 5/6. The antisense transcript was only amplified in samples treated with RNA duplexes containing a biotinylated sense strand. This was seen for T47D (**Fig 24f)** and MCF7 **(Fig 24e**) cells for PR-9 (silencing) and PR-11 (activation). Using primers directed at the PR genomic DNA no direct association of agRNAs with genomic DNA was detected (**Fig 24e-f)**. For other oligonucleotides that do strand invade, agLNAs, this protocol did detect association with genomic DNA (*235*).

Recall that chromatin immunoprecipitation did see argonaute recruitment to the DNA for silencing RNA PR-26 (*240*). However, this is due to the fact that chromatin immunoprecipitation uses formaldehyde crosslinking to chemically bind proteins to nucleic acids. For the biotin pulldown, no crosslinking is used so only the most direct interactions survive pulldown and washing.

For these experiments, the transcript AT-2 was detected by primers spanning exon boundaries. This suggests that the association detected was with spliced transcripts. Since splicing occurs co-transcriptionally (*285-287*), these may still be nascent transcripts but the location of the exons would place them far away from the PR promoter. Alternatively, these transcripts may remain associated with chromatin post-processing and thereby coat the DNA near the promoter for PR, as is seen for the noncoding RNAs Air and Xist (*179, 184*). The interaction detected also may be an interaction that occurs away from the PR promoter and not directly involved in PR regulation but indicative of the fact that the noncoding RNA is recognized by agRNAs in general. However, the fact that nuclei were isolated does suggest that the detected interaction occurs in the nucleus.

These experiments agree well with model 2 (**Fig 12**) and provide evidence that agRNAs do not, in fact, interact directly with genomic DNA. For these experiments, it would be helpful to know more about the nature of the interaction of the antisense transcript and the PR promoter. For this, RNA-FISH or RNA-TRAP assays might be illuminating (*179, 184*).



S. Younger

Figure 24: Biotin pulldown of agRNAs with their molecular targets. (A) Biotinylated agRNAs activate (A and B) and silence (C and D) as well as unmodified RNA duplexes. Activating agRNA PR-11 pulls down the antisense transcript without evidence of a direct interaction with genomic DNA (E). Silencing agRNA PR-9 pulls down the antisense transcript without evidence of a direct interaction with genomic DNA (F).

4.6 agRNA ACTIVATION IS AT THE LEVEL OF TRANSCRIPTION

Although agRNA silencing had been shown to act at the level of transcription by nuclear run-on assay, agRNA activation had not. For this experiment I used chromatin immunoprecipitation of RNA polymerase II (RNAP2) (clone CTD4H8, Millipore 05-623). I measured the relative abundance of RNAP2 at the transcription start site of PR-B.

I transfected T47D cells with either MM or PR-9 agRNAs at 25 nM with RNAi-Max and MCF7 cells with either MM or PR-11 agRNAs at 25 nM. In T47D, silencing of PR was verified by qRT-PCR. The agRNA PR-9 led to marked reduction of RNAP2 three days after transfection (**Fig 25b**). For MCF7 cells, activation was verified by qRT-PCR. The activating agRNA PR-11 led to markedly higher recruitment of RNAP2 to the PR promoter after 3 days (**Fig 25a**). These effects validate that agRNAs regulate at the level of transcription for both silencing and activating agRNAs.

4.7 agRNAS RECRUIT ARGONAUTE PROTEINS TO THE ANTISENSE TRANSCRIPT

Silencing agRNAs involve argonaute proteins. However, this had not been clearly shown for agRNA activation. I collaborated with D. Hardy and B. Janowski to perform chromatin immunoprecipitation of argonaute proteins in T47D and MCF7 cells. For these experiments, a well characterized antibody was acquired from the Z. Mourelatos laboratory (clone 2A8, (*288*)). This antibody is reported to bind and pulldown all four of the human argonaute proteins.

B. Janowski transfected T47D cells with 25 nM of either MM or PR-9 agRNAs. Silencing of PR was verified by western analysis. D. Hardy performed ChIP of argonaute proteins using the Mourelatos anti-Ago antibody. Argonaute protein was found enriched at the PR promoter for agRNA silencing (**Fig 25d**). This data suggests that silencing agRNAs recruit argonaute proteins to the PR promoter.

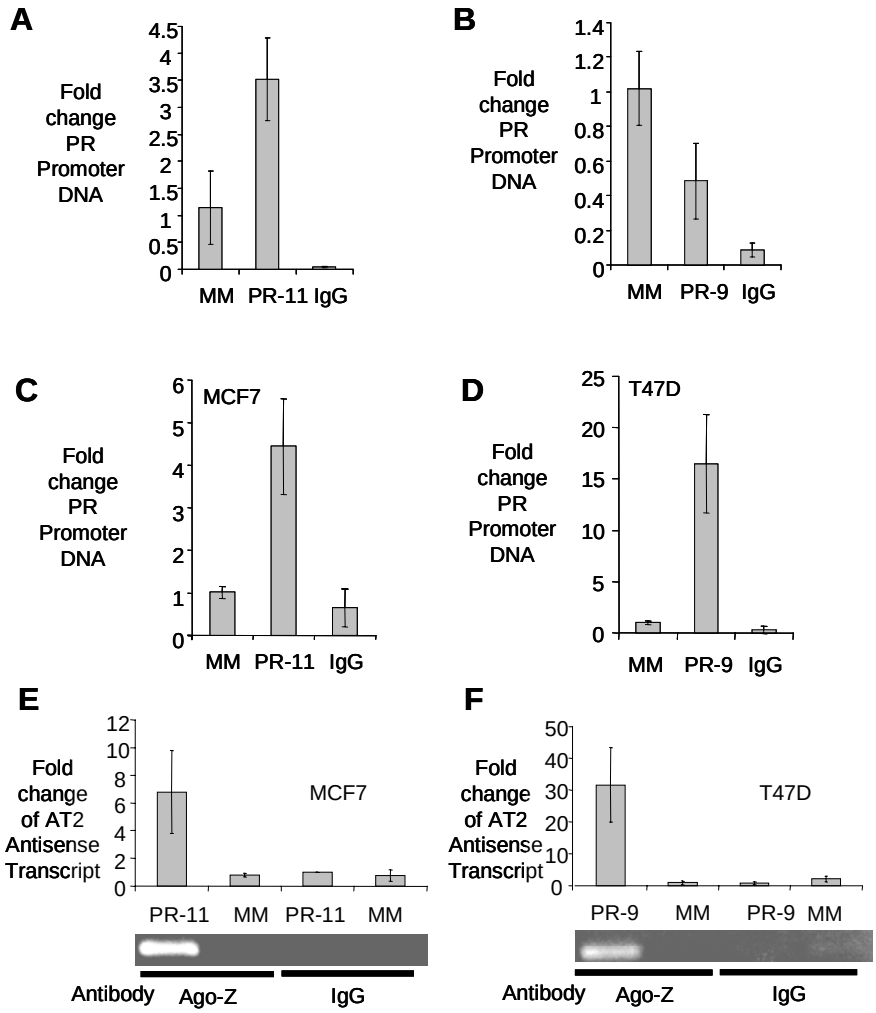


Figure 25: RNAP2 and argonaute immunoprecipitations. Chromatin immunoprecipitation of RNAP2 (CTD4H8) reveals increased RNAP2 for gene activation (A) and reduced RNAP2 for gene silencing (B). Chromatin immunoprecipitation of argonaute proteins (2A8) reveal recruitment of argonaute to the PR promoter for both gene activation (C) and gene silencing (D). RNA immunoprecipitation of argonaute proteins (2A8) reveal that agRNAs recruit argonaute to the antisense transcript both in gene activation (E) and gene silencing (F). Data is qRT-PCR averaged from at least 3 independent immunoprecipitations. Error bars are standard deviations.

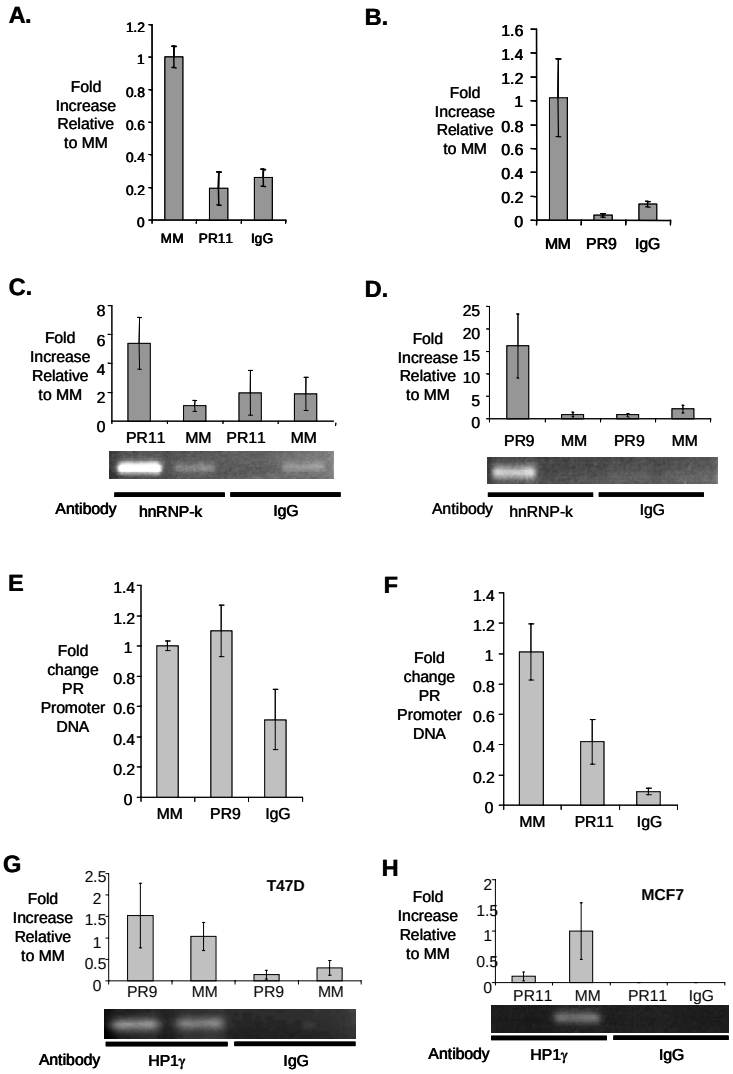


Figure 26: Changes in hnRNP-k and HP1γ. ChIP reveals dismissal of hnRNP-k (3C2) from the PR promoter by activating (A) and silencing (B) agRNAs but RIP reveals recruitment to the antisense transcript by activating (C) and silencing (D) agRNAs. ChIP reveals no change for HP1γ (43S2) in gene silencing (E) but dismissal in gene activation (F). RIP reveals no change in HP1γ interactions with the antisense transcript in gene silencing (G) but also dismissal by gene activation (H). Data is qRT-PCR averaged from at least 3 independent immunoprecipitations. Error bars are standard deviations.

Following this, B. Janowski transfected MCF7 cells with 25 nM of either MM or PR-11 agRNAs. Silencing of PR was verified by western analysis. D. Hardy performed ChIP of argonaute proteins and also saw enrichment at the PR promoter for agRNA activation **(Fig 25c**). This data suggest that activating agRNAs recruit argonaute proteins to the PR promoter

I then collaborated with N. Nguyen to test the association of argonaute proteins with the antisense transcripts. For this I used RNA immunoprecipitation (RIP). N. Nguyen and I transfected and harvested T47D cells with 25 nM either of MM or PR-9 agRNAs. I validated silencing by qRT-PCR. After immunoprecipitation with antibodies targeting argonaute, I performed RT-PCR to amplify the antisense transcript. The antisense transcript was detected in samples treated with PR-9 but not with MM or in IgG negative control samples (**Fig 25f**). This suggests that silencing agRNAs recruit argonaute proteins to the antisense transcript.

N. Nguyen and I then transfected and harvested MCF7 cells with 25 nM of either MM or PR-11 agRNAs. N. Nguyen and I validated activation by qRT-PCR. After immunoprecipitation with anti-Ago antibodies, I performed RT-PCR to amplify the antisense transcript. The antisense transcript was detected in samples treated with PR-11 but not with MM or in IgG negative control samples (**Fig 25e**). This suggests that activating agRNAs recruit argonaute proteins to interact with the antisense transcript.

For these experiments I reported qRT-PCR measurements of the relative abundance of the antisense transcript. However, recovery of RNA for this protocol was very inefficient with C_T values in the high 30's, well outside the linear range for these primers. C_T values for MM treated or IgG samples were in the low 40's. However, by gel, no product was observed for these samples, only primer dimers. For this reason I do not believe that qRT-PCR is an acceptable way to measure RIP recovery of RNA. Instead, in this case the agarose gel appears more reliable. Using agarose gel, the only determines whether or not argonaute proteins associate with the transcript and say nothing about the relative abundance of association.

4.8 agRNAS REORGANIZE INTERACTIONS FOR HNRNP-K AND HP1γ

B. Janowski hypothesized that the protein hnRNP-K may have a role in regulation of PR. hnRNP-K is a protein known to bind both DNA and RNA and can act as a transcription factor. Putative hnRNP-K binding sites are in the PR promoter and B. Janowski gathered preliminary evidence that hnRNP-K regulates PR expression in MCF7 cells (B. Janowski unpublished data).

B. Janowski and D. Hardy collaborated to test by chromatin immunoprecipitation if hnRNP-K recruitment to the PR promoter is changed by treatment with agRNAs. B. Janowski transfected T47D cells with 25 nM of either MM or PR-9 agRNAs. RNA induced silencing was measured by western. D. Hardy performed ChIP with anti-hnRNP-K antibody (clone 3C2, Sigma Aldrich, R8903). He found that in T47D cells, hnRNP-K is dismissed from the PR promoter upon transfection with agRNA PR-9 (**Fig 26b**).

B. Janowski then transfected MCF7 cells with either 25 nM MM or PR-11 agRNAs. Gene activation was verified by western. D. Hardy performed ChIP with anti-hnRNP-K antibody (3C2) and found that also in MCF7 cells, hnRNP-K is dismissed from the PR promoter upon transfection with agRNA PR-11 (**Fig 26a**).

N. Nguyen and I collaborated to determine if hnRNP-K also changed in relation to association with the antisense transcript. N. Nguyen transfected T47D cells with either 25 nM MM or PR-9. I validated gene silencing by qRT-PCR. Contrary to the ChIP results, by RIP N. Nguyen and I found hnRNP-K recruited to the antisense transcript by transfection with silencing agRNA PR-9 (**Fig 26d**).

N. Nguyen then transfected MCF7 cells with either 25 nM MM or PR-11 agRNAs. I validated gene activation by qRT-PCR. Again, RIP showed association of hnRNP-K with the antisense transcript in PR-11 treated cells and not in MM (**Fig 26c**).

Next, I hypothesized that the protein HP1 may have a role in mammalian agRNA activity. HP1 has a role in TGS in *S. pombe* and is recruited by the RITS complex. In human there are 3 HP1 proteins – α, β, and γ. HP1α and HP1β have been shown to associate with constituitive heterochromatin, but HP1γ has been observed to associate

with both heterochromatin and euchromatin. For this reason I decided to study the role of HP1γ in the agRNA activity.

I treated T47D cells with either 25 nM MM or PR-9 agRNAs. I validated gene silencing by qRT-PCR. I performed chromatin immunoprecipitation using anti-HP1γ antibodies (clone 43S2, Millipore 05-690). I found that HP1γ association was unchanged by treatment with PR-9 silencing agRNAs (**Fig 26e**).

I then treated MCF7 cells with either 25 nM MM or PR-11 agRNAs. I validated gene activation by qRT-PCR. I performed chromatin immunoprecipitation and found HP1γ to be dismissed from the PR promoter by activating agRNA PR-11 (**Fig 26f**).

I then tested if HP1γ associated with the antisense transcript and if this was changed by treatment with agRNA. I treated T47D cells with either 25 nM MM or PR-9 agRNAs and perform RNA immunoprecipitation. I found that HP1γ does associate with the antisense transcript in T47D cells but this association is unaltered by treatment with agRNA PR-9 (**Fig 26g**).

I then treated MCF7 cells with either 25 nM MM or PR-11 agRNAs and performed RNA immunoprecipitation. I validated gene activation by qRT-PCR. I found that HP1γ associates with the antisense transcript but is dismissed by treatment with agRNA PR-11 (**Fig 26h**).

The hope in these experiments was to uncover a potential protein partner for argonaute that might have a role in mediating agRNA activity in general. The results from these experiments do not point toward a role for hnRNP-K or HP1γ as a protein partner for argonaute because these proteins do not follow argonaute in its recruitment to the promoter DNA and to the antisense transcript by agRNAs.

However, these results do point towards a restructuring of protein interactions on the PR promoter induced by agRNA targeting. A large question remains about why a disconnect is observed between associations with the RNA and the DNA. Here again it would be helpful to know more about the nature of the interaction of the antisense RNA with the chromatin at the PR gene locus and how that interaction might be altered by targeting with agRNAs.

4.9 IS THE ANTISENSE TRANSCRIPT NONCODING?

The antisense transcript is capped, spliced, and polyadenylated just as a coding mRNA would be, but this does not mean that it is coding (*289*). For an empirical approach, it would be helpful to show if the antisense transcript escapes the nucleus and whether it associates with ribosomes. What can be appreciated is that its coding potential is very low. The transcripts AT-1, AT-2 MCF7, and AT-2 T47D have a higher density of stop codons than is typically found in protein coding genes. This results in most of the positive reading frames being quite short, translating to short peptides only around 30 amino acids (ORF Finder, http://www.ncbi.nlm.nih.gov/projects/gorf). Aligning these potential reads using either Blastp or Blastx reveal no conserved domains or significant alignment with human proteins or nonhuman proteins (**Fig 27**).

Furthermore, the noncoding RNAs in general are poorly conserved. Other than the region overlapping the PR gene, especially the protein coding region, a verbebrate Multiz alignment and conservation analysis across 44 species shows no conservation of transcripts AT-1, AT-2 MCF7, or AT-2 T47D throughout mammalian (http://genome.ucsc.edu). Exons 3, 4, and 5 of the AT-2 transcripts are contained in repeat elements that are not conserved outside of primates. The Riken Fantom3 database does report the existence of an unspliced antisense transcript overlapping the 5' end of the PR gene in mouse, but other than the region overlapping the PR gene, this transcript contains almost no sequence similarity to the human AT-1 (**Fig 28**).

However, the antisense transcript may have some conserved function. The fact that mouse has an antisense transcript overlapping the PR promoter suggests that an antisense transcript may be necessary for normal progesterone receptor regulation. Both transcripts AT-1 and AT-2 have SP1 and estrogen receptor binding sites at or upstream of their transcription start sites. Work by Gingeras and colleagues reveals a short RNA detected in exon 2 of the spliced AT-2 transcript (**Fig 28**). Finally, it seems unlikely that cells would expend all of the energy required to transcribe such a long transcript and splice and polyadenylate it by accident. What the role of antisense transcripts play in regulating gene promoter activity remains to be determined.

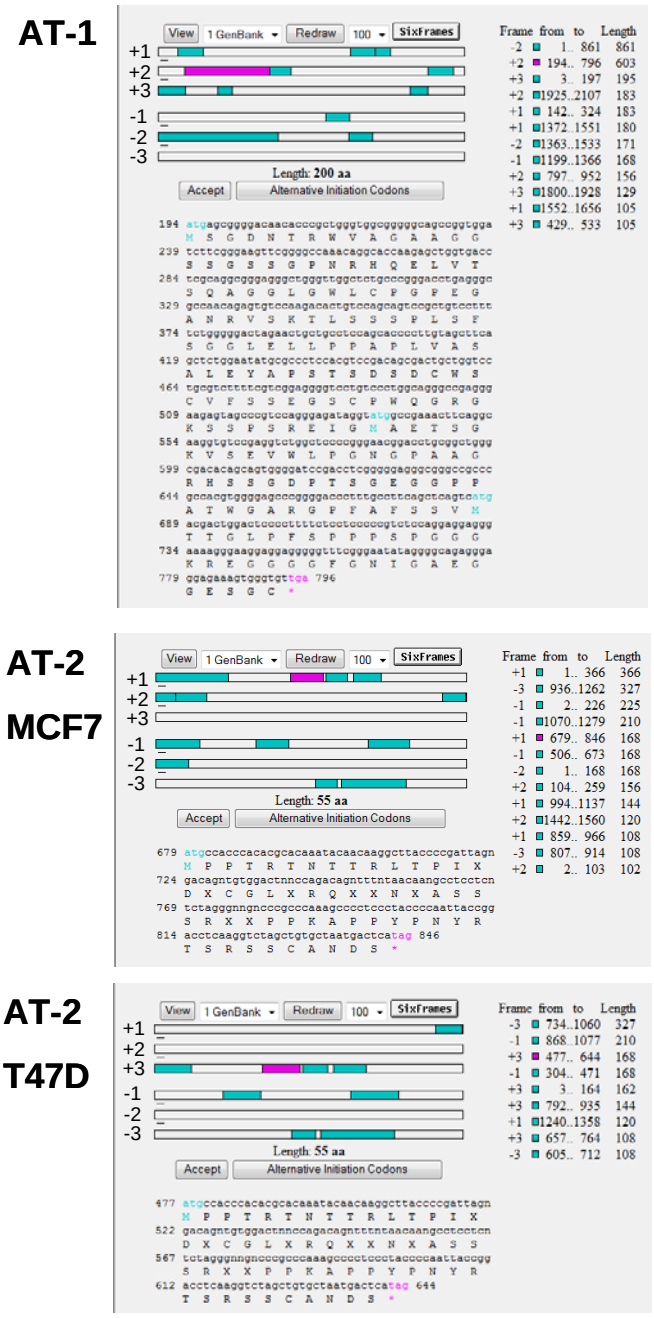


Figure 27: ORF finder shows little coding potential for noncoding RNAs. Most open reading frames in the noncoding RNA transcripts are less than 40 amino acids long and have no conservation.

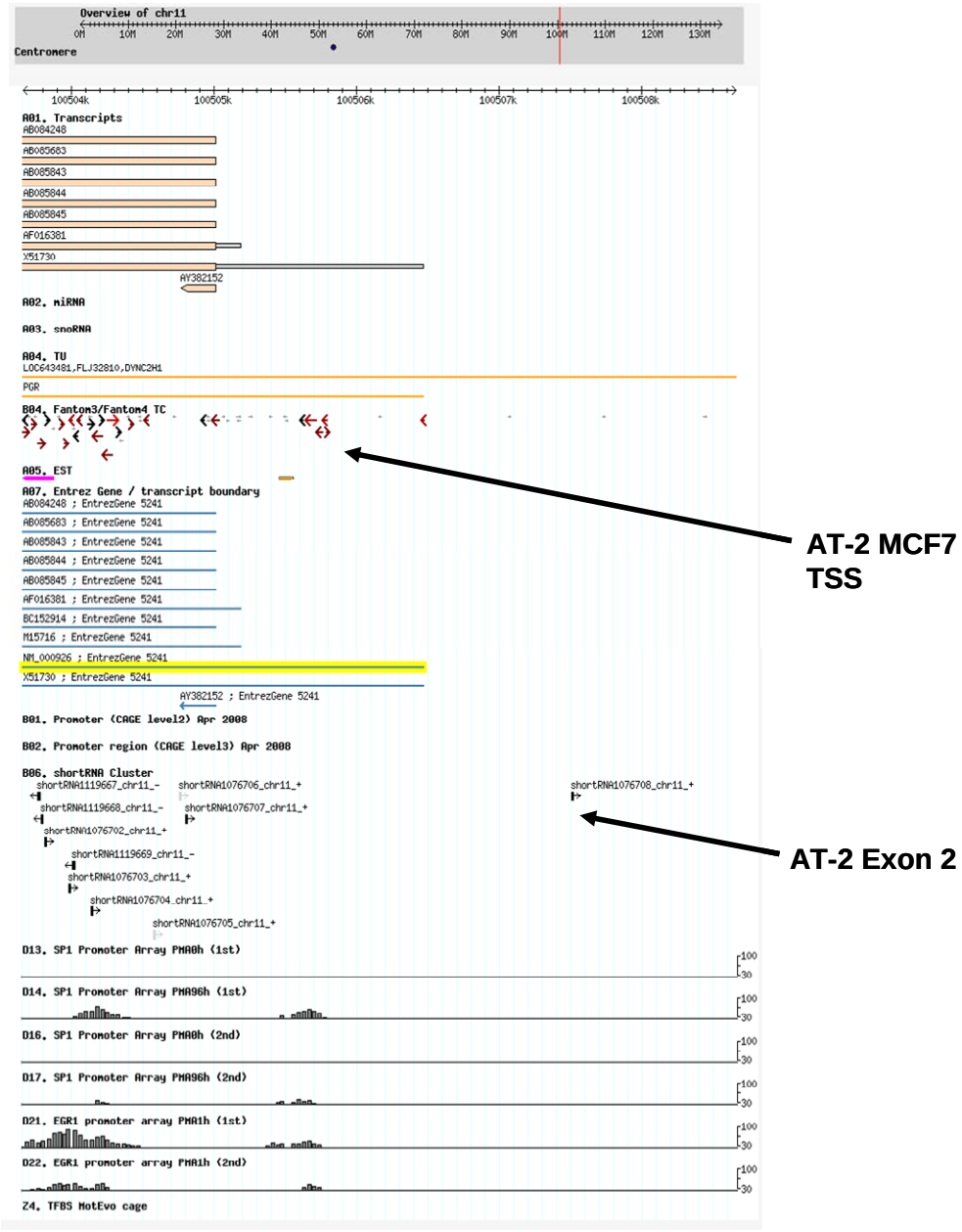


Figure 28: FANTOM4 data on PR promoter. The Fantom4 database has uncovered several sense and antisense transcription start sites by CAGE analysis at the 5' end of the PR gene (NM_000926). Short RNA sequencing also reveals the existence of short RNAs from this region and ChIP analysis notes several SP1 and estrogen receptor binding sites in the region.

4.10 OTHER AGRNA TARGETS HAVE TRANSCRIPTS IN THEIR PROMOTERS

To see how general promoter associated RNAs are I used RT-PCR to determine if promoter RNAs are associated with other genes targeted by agRNAs. I chose the genes Huntingtin (HTT) and p53. For each gene I designed primers that spanned downstream of the transcription start site and into the promoter. For HTT, I detected RNA transcripts as far upstream as -260 nts (**Fig 29**). For p53, I detected RNA transcripts as far upstream as -310 nts (**Fig 29**). For p53, a paper was published later suggesting that an antisense transcript overlaps the promoter of p53, named Wrap53. This transcript is the result of an alternative 5' exon for a protein coding gene, Wdr79 (*290*).

In collaboration with S. Younger, I characterized noncoding transcripts in the promoter of the gene Troap. S. Younger targeted this gene with antigene RNA and achieved transcriptional silencing. Biotin pulldown assay revealed his target to be a sense transcript (S. Younger unpublished results). I first determined the transcription start site for this gene and found it to lay some 36 nucleotides upstream of the refseq annotated transcription start site. As with PR, a distribution of TSSs is seen (**Fig 30, top**).

Next I designed primers to several sites with the 1000 nucleotides upstream of the transcription start site for Troap and performed 5' RACE in MCF7 and T47D cells. I found no sense transcripts in this region but did find two antisense transcripts (**Fig 30, bottom**). I then moved my primers into an upstream EST and found a sense transcript overlapping the first three exons of Troap and ending in intron 3. Interestingly, the agRNA target site lies within an intron of this transcript (**Fig 31**). However, as a nascent transcript, this transcript would not be spliced and may pose as a target for agRNAs. These results pose two interesting possibilities: 1) agRNAs may be able to target sense transcripts, and 2) agRNAs may bind pre-spliced RNA.

Recently, M. Matsui in the Corey lab has uncovered activating agRNAs targeting the gene LDLR. He used 5' and 3' RACE to identify an RNA transcript that might be the target for his agRNAs. He uncovered a long unspliced antisense transcript overlapping the LDLR promoter (M. Matsui unpublished).

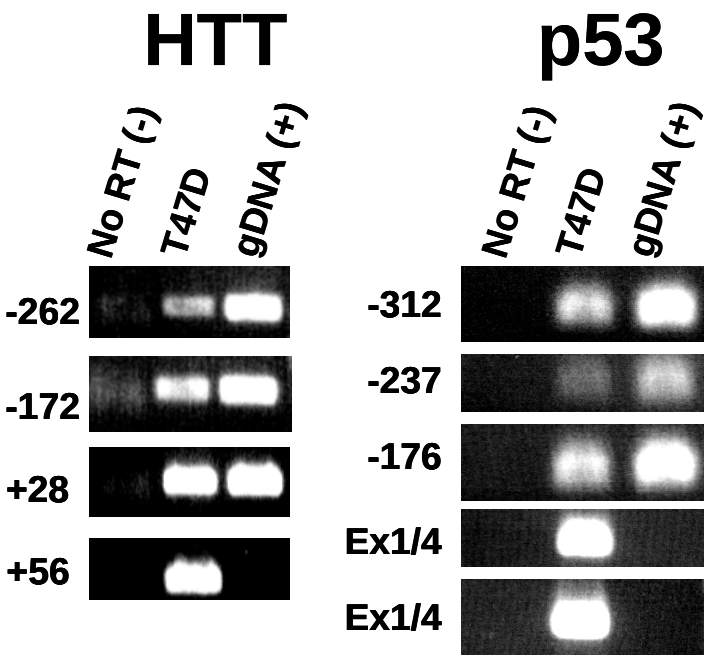


Figure 29: RNA transcripts are expressed from the promoters Huntingtin and p53. RT-PCR using primers targeting downstream and upstream of the transcription start sites for huntingtin and p53 reveal RNA transcripts expression from these gene's promoters. No RT controls show that detected transcripts are not due to contaminating genomic DNA. Lower primer sets span exon boundaries and therefore do not amplify genomic DNA in the positive control lane.

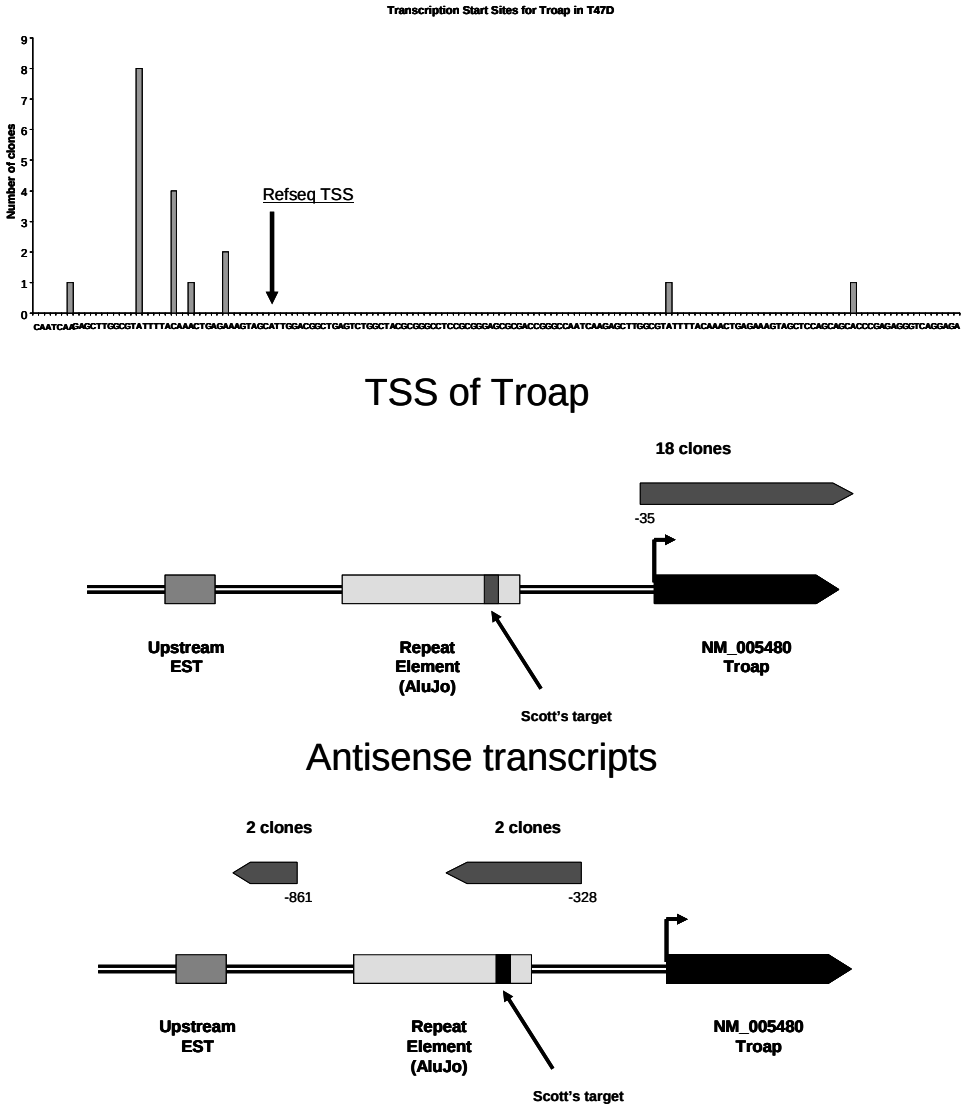


Figure 30: The transcription start site for Troap. (Top and middle) A distribution of transcription start sites is seen for the gene Troap as was seen for PR-B (n=18). The major transcription start site is 36 nucleotides upstream of the annotated TSS. (Bottom) Two antisense transcript were found in the Troap promoter.

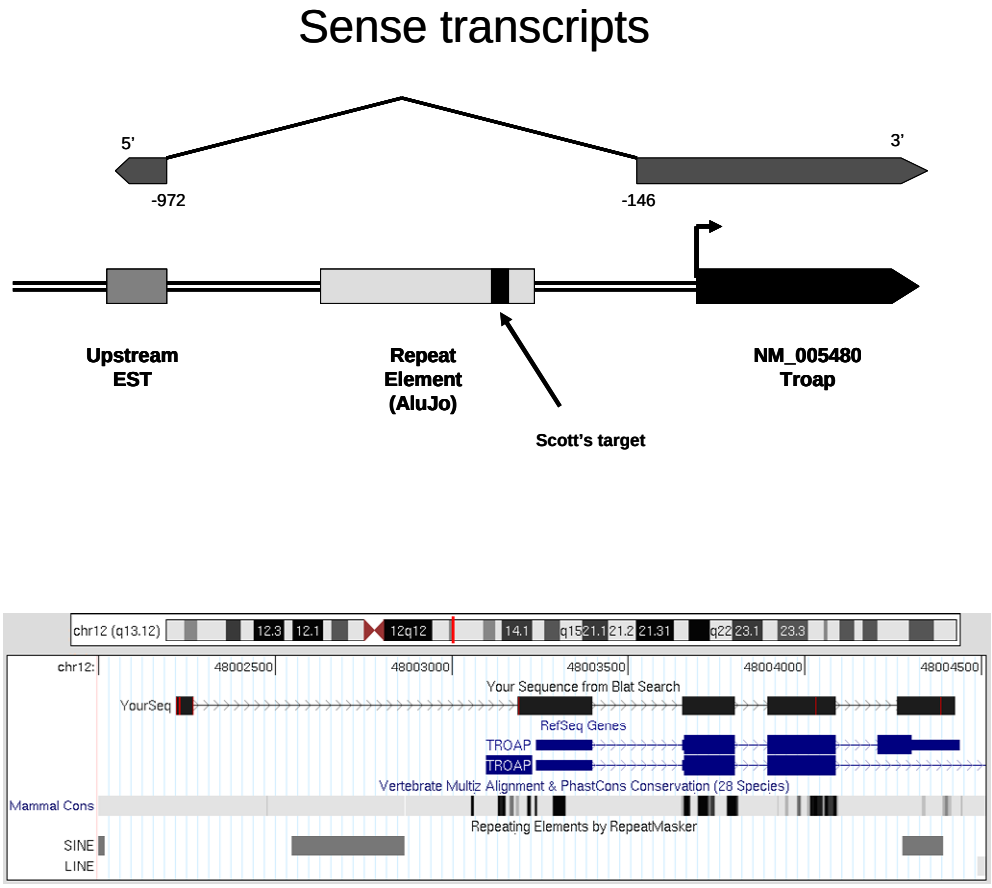


Figure 31: A sense transcript overlaps the Troap promoter. The only sense transcript overlapping the Troap promoter found was a spliced transcript whose intron overlaps the Troap promoter. This transcript is spliced with five exons, three of which overlap exons from the Troap gene. A truncated Troap transcript has been deposited in Refseq that aligns with 4 exons of the Troap sense transcript. This transcript is annotated as having coding potential but to date that has not been demonstrated.

Although it has not been targeted with agRNAs, I have analyzed the promoter of the gene CCND1. In mouse, noncoding RNAs in the promoter of CCND1 have been shown to recruit the protein TLS, inhibit histone acetylation, and silence transcription (*197*). These transcripts have been targeted with siRNAs which leads to gene activation. The authors interpret this to be a reversal of TLS recruitment by the noncoding RNAs but it may also be due to an agRNA related gene activation.

I used 5' RACE to identify the transcription start site for CCND1 in human MCF7 cells and found two transcription start sites at +52 and +109. Again this result suggests a distribution of start sites for the gene CCND1. I then used primers targeting the most prominent TLS binding site at -1000 nucleotides. This is also the site that was targeted with an siRNA in mouse to induce gene activation. I did not find evidence of a sense transcript but I did find an antisense transcript with a transcription start site at -603.

4.11 IS THE ANTISENSE TRANSCRIPT AT-2 THE TARGET OF agRNAS?

The evidence presented here suggests that for PR agRNAs interact directly with the antisense transcript and not with genomic DNA. Argonaute proteins seem to have two fold interactions with both the RNA and the DNA. The nature of the interaction of argonaute with DNA is still unresolved. Also the nature of the interaction between the antisense transcript and the PR promoter remains a question.

The data in this study in no way precludes the possibility of agRNAs targeting a sense transcript (*291*). The gapmer data suggests that the noncoding RNA's role is limited to being a substrate rather than being a regulatory mechanism in itself, unlike the mechanism reported by Morris et al (*284*). Data from S. Younger suggests that a sense transcript can be a target. Further data from X. Yue also suggests a sense transcript as a target for agRNAs. This will be discussed in the next chapter.

If an RNA:DNA hybrid between agRNAs and chromosomal DNA is not formed, then the nature of the interaction between argonaute and chromosomal DNA is likely through protein:RNA, protein:protein, and protein:DNA interactions. In the case of RITS, the complex member Chp1 binds methylated histone tails and thereby anchors argonaute

to the DNA (*41*). The proteins HP1 and the Polycomb complex also bind methylated histone tails and may serve an analogous role to Chp1 from *S. pombe*. However, there is no strong evidence of such an interaction yet.

4.12 THE TRANSCRIPT AT-2 MAY ENCODE A MICRORNA

One of the first observations I made concerning the antisense transcript AT-2 is that three of the exons are contained in repeat elements and have complementarity to many other regions of the genome (**Fig 32**). This would be expected of repeat elements but I further noticed that many of these hits are with protein coding genes. I collaborated with Scott to make alignments of exons 3, 4, and 5 with all Refseq mRNAs using a very stringent criteria of at least 15 nucleotides complementarity. Exon 3 and 5 did not show significant complementarity to protein coding genes using this criteria. However, exon 4 aligned with more than 550 protein coding genes (**Fig 32**). Interestingly, 71% of all hits aligned within the last third of the mRNA sequence, suggesting a bias towards 3' ends or 3' UTRs (**Fig 33, top**).

I measured the frequency of hits aligning in exon 4 by counting the number of times and alignment began on a particular nucleotide in the exon. I found a strong bias for alignments to be with a sequence on the 5' of exon 4 (**Fig 33, middle**). I then used mFOLD to make secondary structure predictions for the transcript with exons 3, 4, 5. Even using several different large and small truncations of the AT-2 transcript always gave the region around the 5' end of exon 4 forming a hairpin, with the sequence aligning with the 3' UTRs of mRNAs lying on one side. Alignments also tended to preserve complementarity in nucleotides 1 through 7 (**Fig 33, bottom**).

All of these facts suggest that the transcript AT-2 may encode a microRNA. However, attempts to verify this by checking in cells treated with antisense gapmers have yield conflicting and inconsistent results. This may be due to general toxicity associated with gapmers. A better approach to these experiments may be to perform a northern to try to detect the mature miRNA being produced and then developing an antagomir approach.

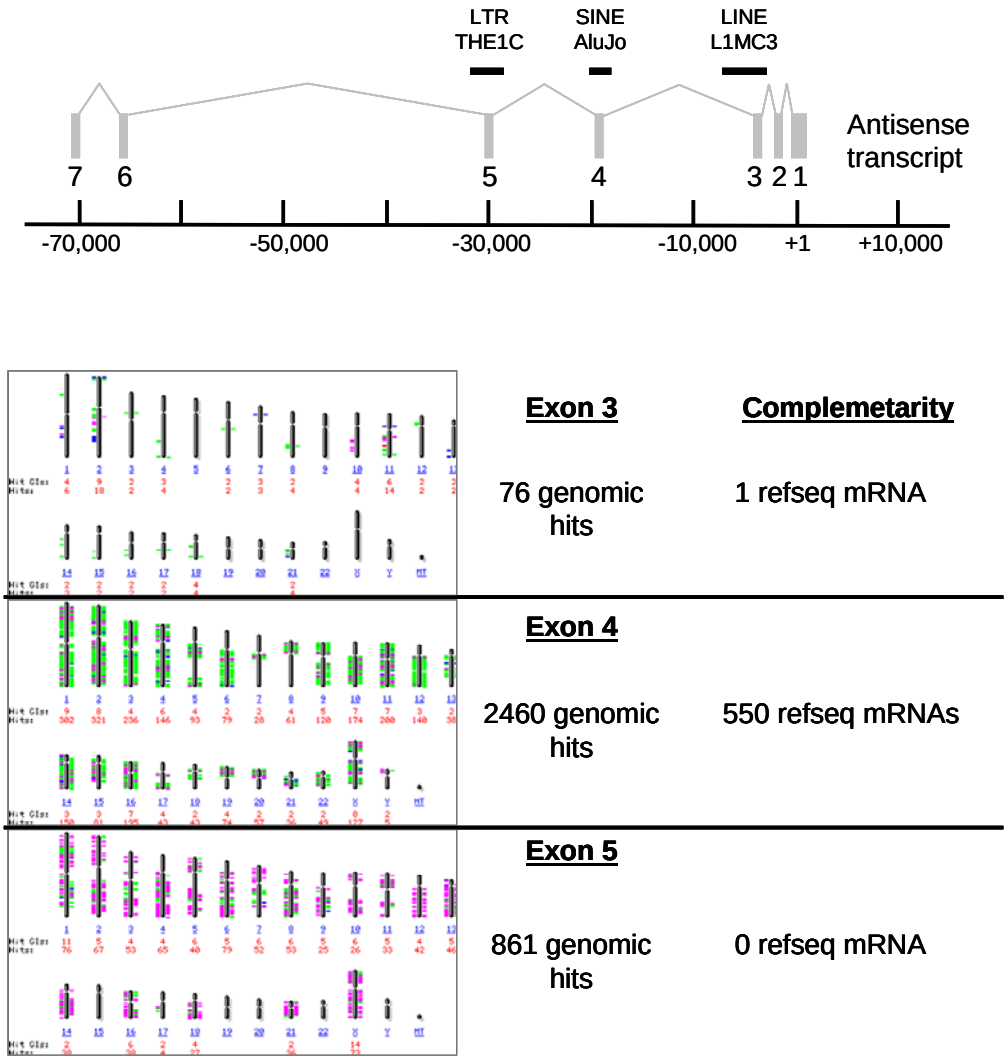


Figure 32: The transcript AT-2 contains repeat elements. (Top) The transcript AT-2 contains three repeat elements that maintain a high degree of complementarity to other regions of the human genome. (Bottom) Using a high stringency alignment, exons 3 through 5 match many sites in the human genome and exon 4 has many matches with annotated mRNAs.

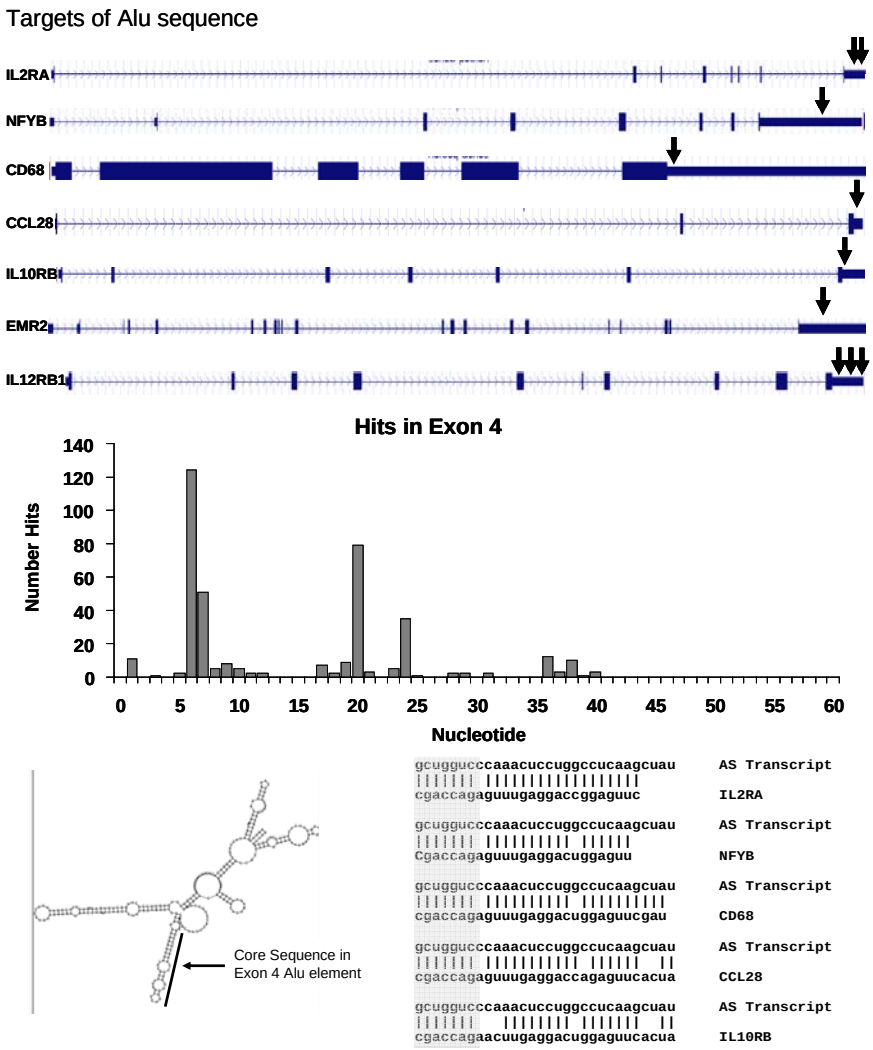


Figure 33: Exon 4 of AT-2 may encode a miRNA. (Top) Many of the mRNAs complementary to Exon 4 of AT-2 have hits in their 3'-UTRs. (Middle) Aligning regions of complementarity in Exon 4 by their 5' ends reveals a preference for a region towards the 5' end of exon 4 to align with other mRNAs. (Bottom) mFOLD structure predictions of AT-2 reveal the core sequence on Exon 4 to lie on a hairpin structure. Also, alignment of Exon 4 with mRNAs reveal a bias towards a "seed sequence" like match with nucleotides 1 through 7 not containing mismatches.

4.13 THE ROLE FOR NONCODING RNAS IN agRNA ACTIVITY

In 2007, a paper appeared from the Morris group claiming that an RNA transcript existed as the substrate for agRNA mediated TGS (*292*). Unfortunately, the question remains unclear based on the data presented. It is difficult to determine for this paper if the sequence studied acts through an agRNA mechanism since 1) a single stranded RNA was used rather than double stranded which is known to be ineffectively loaded into argonaute *in vivo*, 2) the knockdown of the gene was never more that 50% with the error bars overlapping the expression levels of the control, 3) no evidence of knockdown at the level of transcription was shown, and 4) the paper showed that the RNA target was an mRNA variant further raising the question if the measured effect was related to transcriptional or post-transcriptional regulation.

The evidence given in this paper that this promoter RNA was involved was 1) an interaction between the RNA and the biotinylated single stranded RNA, and 2) reversal of the small levels of gene silencing by silencing the targeted RNA with an antisense oligonucleotide. Again, these two lines of evidence were highly questionable. For the biotin pulldown there was no negative control, single stranded RNAs were used, and no evidence was shown the biotinylated, single stranded RNAs could actually silence gene expression. For the antisense oligonucleotide experiments, problems included that reversal was minimal due to minimal levels of knockdown, the control oligonucleotide also caused knockdown but the experiments were normalized so as to mask this effect leaving the reversal experiment uninterpretable, and the antisense oligonucleotide used was synthesized with older, inferior chemistry of phosphothioate which is known to have problems with toxicity in cell culture (*14*).

In 2008, another Morris et al. paper concluded that gene activation is the result of cleavage of antisense transcripts. This paper postulates the existence of an antisense transcript overlapping the whole of the p21 gene based on the claim of reverse transcription using the strands of an siRNA duplex for primers. Since reverse transcriptase does not extend RNA:RNA duplexes, one can only assume that this is a

misrepresentation and that DNA was used rather than RNA (*87*). The premise for this experiment is that a primer designed to reverse transcribe RNA in the p21 promoter actually nonspecifically reverse transcribed the p21 mRNA and an antisense transcript overlapping the p21 mRNA. This is taken to support the notion that the promoter targeting duplex actually nonspecifically targets within the p21 coding region, recognizing an antisense transcript within the p21 coding region and cleaving it, while the other strand of the duplex is assumed to do nothing to the mRNA. This off-target recognition is postulated based on 12 out of 19 nucleotides being complementary. These 12 nucleotides of complementarity are clustered on the 3' end of the duplex whereas it is well established that mismatched miRNA-like interactions strongly depend on seed sequence complementarity at the 5' end of the duplex (*293*).

Morris et al. perform quantitative realtime PCR on the nonspecifically amplified product and conclude that the off-target binding of the duplex RNA leads to upregulation of p21 mRNA and downregulation of the antisense transcript. Morris et al did not show evidence that this nonspecific strand-specific qRT-PCR amplifies linearly. Morris et al show an illegible nuclear run-on blot marred with nonspecific blotches and shadows to argue that the activation is at the level of transcription since a 1.27 fold increase is seen ($p=0.0841$).

A 46% knockdown of Ago2 ($p= 0.35$) leads to a switch of the promoter targeting duplex from acting as a p21 activator to a p21 silencer. The large error bars for these experiments resulting in large p-values is especially concerning since they do not represent three independent experiments. Instead, cDNA from three transfections was pooled into a single qRT-PCR experiment and the error bars represent the technical error between triplicate qRT-PCR measurements. This means that the large error is not due to biological differences in cells or transfections but simply from pipetting error and/or possibly inefficient and inconsistent PCR amplification between technical replicates of the same qRT-PCR experiment.

This paper continues to make many claims, often based on data where the standard deviations are larger than the average. Morris et al never explain how in their model this duplex nonspecifically recognizes an antisense transcript and cleaves it but

when it nonspecifically recognizes the mRNA, no cleavage occurs. In short, in the case of PR activating RNAs that work through sequence-specific recognition of an antisense transcript overlapping the PR promoter, there is no evidence that cleavage of the antisense transcript occurs during gene activation and cleavage of the antisense transcript with antisense gapmers has no effect on PR expression. Furthermore, more recently, X. Yue has shown gene activation by targeting a sense transcript (Chapter 5). In this case again, no cleavage of the noncoding RNA is observed.

CHAPTER FIVE: Antigene RNAs Targeting the 3' Terminus

5.1 PROGESTERONE RECEPTOR 3' UTR

In 2006, X. Yue in the Corey lab initiated a comprehensive study on targeting the 3' UTR of progesterone receptor with duplex RNAs. In the course of this study, it was necessary to characterize the 3' UTR of Progesterone Receptor. Until 2008, the 3' UTR remained as poorly annotated as the 5' UTR. Literature existed suggesting that the PR mRNA may in fact be much longer than the 4500 nucleotide long annotated transcript (NM_000926.3) (**Fig 35a**) (*281*). Since 3' RACE depends on PCR, it was possible that the 3' UTR of PR is too long to amplify using RACE. For this reason, X. Yue probed the possible 3' UTR of PR using a long RT-PCR experiment.

X. Yue designed primers that amplified from exon 7 to some 6000 nucleotides downstream of the stop codon in exon 8 and amplified an RNA transcript. Primers designed to amplify a longer transcript did not produce product. However, this was more likely to be due to the length of the amplicon than to be due to the lack of an mRNA species. X. Yue designed primer to amplify from 3000 nucleotides downstream of the stop codon to approximately 10000 nucleotides downstream of the stop codon and was also able to amplify product. This suggested the possibility that the PR 3' UTR may be some 10000 nucleotides long, whereas it had been annotated to be 800 nucleotides long.

Following this experiment, X. Yue designed a Northern experiment to determine where the PR gene ended. Papers had suggested that the PR mRNA may be more than 11000 nucleotides long and this agreed well with X. Yue's RT-PCR data but this could also be explained by internal splice variants such as inclusion of an intron. To differentiate these possibilities, X. Yue designed a Northern probes to different putative PR mRNAs (**Fig 35b**).

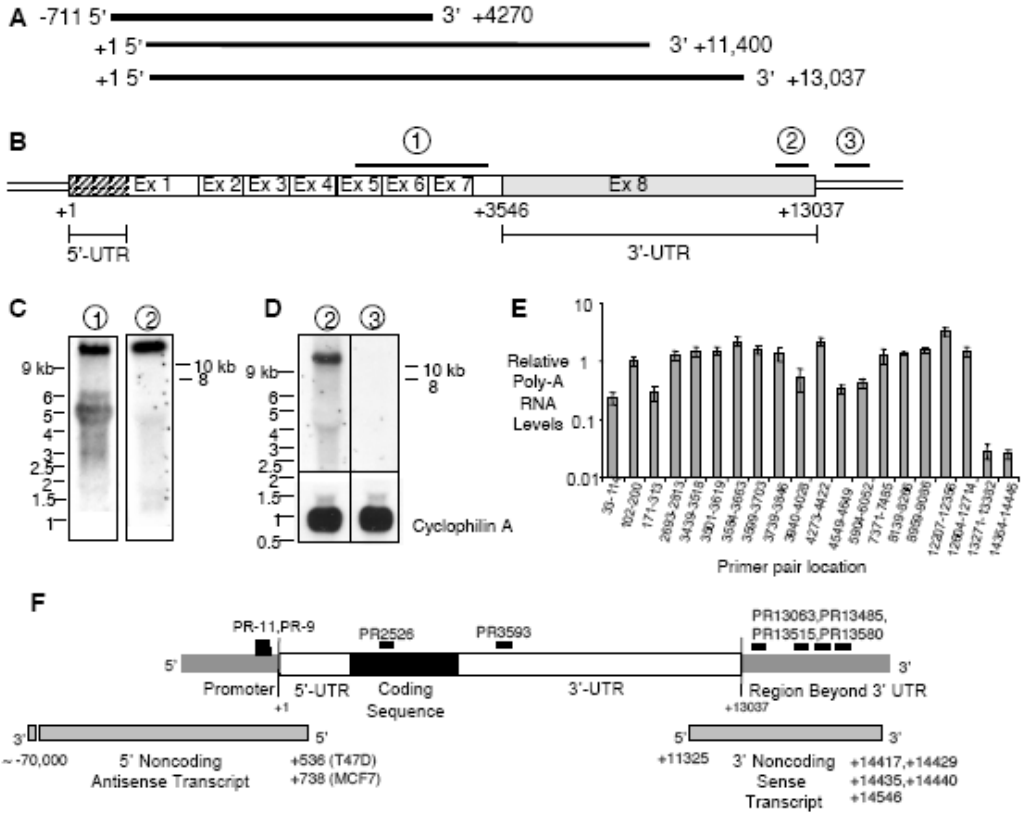


Figure 35: Characterization of PR mRNA. (**A**) Differing annotations of PR mRNA. Top: pre-2008 GenBank (NM_000926.3); Middle: Predicted largest transcript based on Northern analysis in published reports; Bottom: Current GenBank (NM_000926.4). (**B**) Schematic of PR mRNA predicted by GenBank and locations of probes for Northern analysis. (**C**) Northern analysis of PR mRNA comparing results using probes that detect PR mRNA (probe 1) or targeting the 3' termini of PR mRNA (probe 2), (**D**) Northern analysis of PR mRNA using a probe 2 or a probe immediately downstream of the potential mRNA terminus (probe 3). (**E**) qPCR showing levels of poly-A RNA in T47D cells detected from the PR transcription start site past the most downstream annotated terminus of PR mRNA. Notation indicates target region for PCR primers. (**f**) Location of target sequences for duplex RNAs relative to PR mRNA. The 5' and 3' noncoding transcripts that overlap the transcription start site and polyadenylation site are shown.

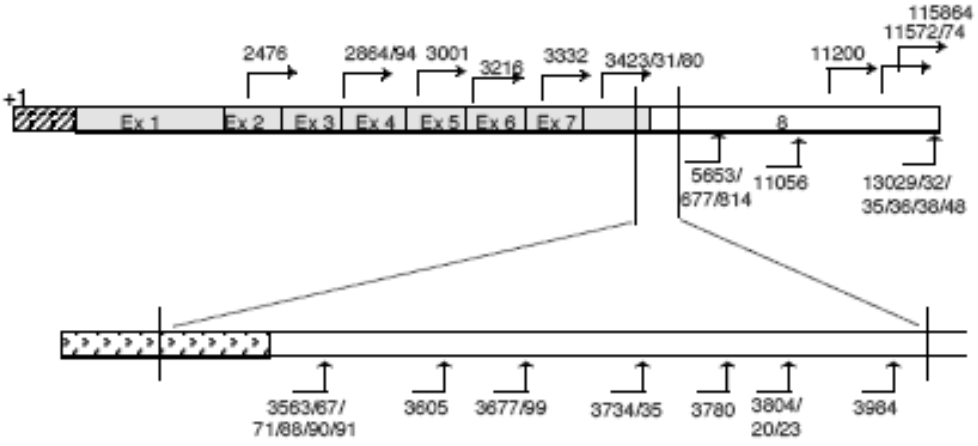


Figure 36: Complex transcription across the PR gene. Transcription start sites and termination sites overlapping the 3' UTR identified by 5' and 3' RACE. Right facing arrowheads indicate transcription start sites. Upward facing arrowheads indicate termination sites. The nucleotides of the start or terminus are indicated.

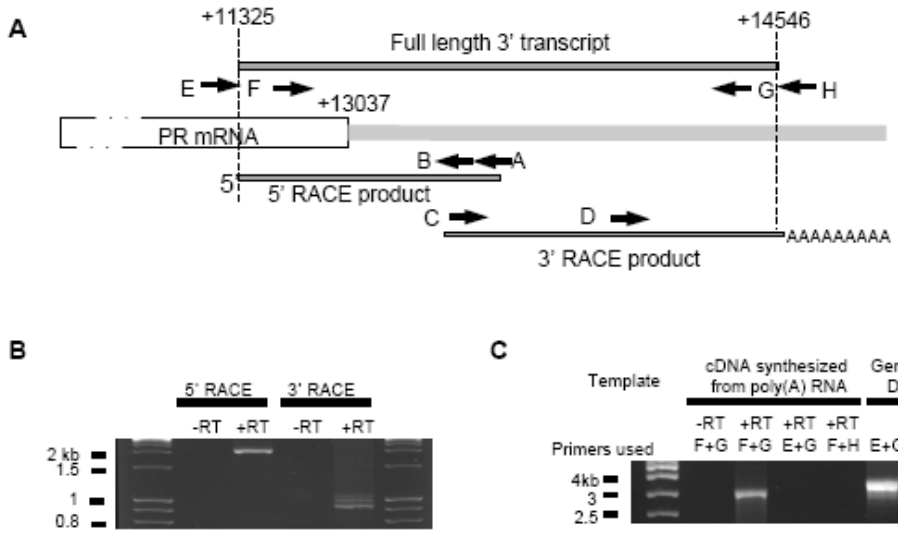


Figure 37: Characterization 3' noncoding transcript. (**A**) Location of RACE (A,B,C, and D) or RT-PCR (E,F,G, and H) primers relative to PR mRNA and the 3' noncoding transcript. (**B**) Agarose gel analysis of RACE products. Total RNA used in RACE was treated with DNase prior to reverse transcription and –RT assay was used as negative control. +RT: Reverse transcriptase added. -RT: No reverse transcriptase added. (**C**). Agarose gel analysis of RT-PCR amplification using primers E, F, G, and H as shown. Poly(A) RNA was DNase-treated prior to reverse transcription and –RT assay was used as negative control. +RT: Reverse transcriptase added. -RT: No reverse transcriptase added. Amplification of genomic DNA was included as a postive control for primer function.

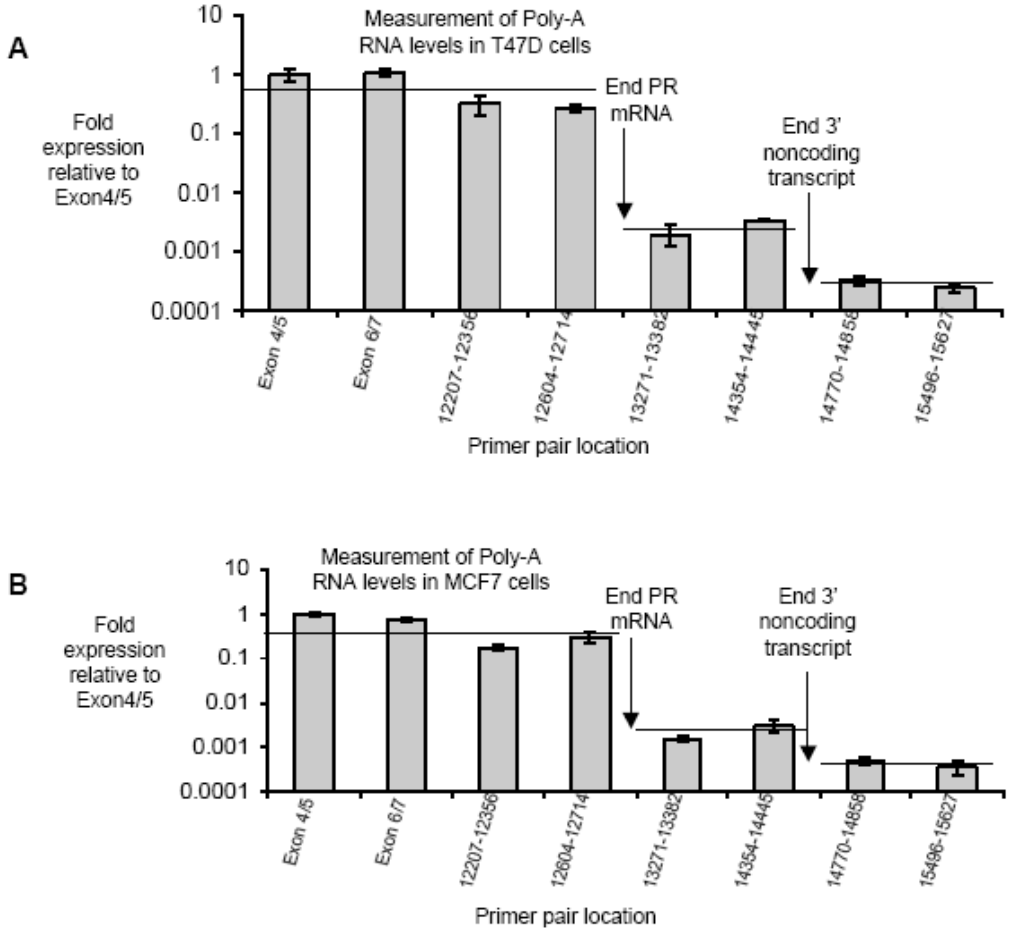


Figure 38: Measurement of poly-A RNA levels. Poly A RNA levels were measured in (**A**)T47D and (**B**)MCF7 cells using different primer sets recognizing regions before the 3' terminus (+13,037) of PR mRNA, after the 3' end of PR mRNA, and downstream from the 3' terminus of the 3' noncoding transcript. qPCR demonstrates that levels of poly-A RNA drop sharply past the 3' end of PR mRNA. RNA levels then drop sharply again for primers complementary to targets past the 3' end of noncoding RNA. Primer set Exon4/5 targets the boundary of exon 4 and 5 in PR mRNA. Primer set Exon6/7 targets the boundary of exon 6 and 7 in PR mRNA.

Previous Northern experiments in our lab had detected a longer transcript but it appeared that the majority of transcripts were on the order of 3000 to 5000 nucleotides long, consistent with a short 3' UTR. However, very long mRNA molecules have difficulty transferring, which introduces bias to the experiment. For this reason, X. Yue optimized a protocol to partially digest mRNA after electrophoresis using NaOH to facilitate more efficient transfer to the nitrocellulose membrane. Using this protocol, it became clear that the major PR transcript was indeed longer than 11000 nucleotides (**Fig 35c**). X. Yue then designed a Northern probe complementary to an EST situated some 10000 nucleotides downstream of the stop codon. This probe also detected a band of the same molecular weight as PR mRNA. X. Yue designed a probe to target beyond 10000 nucleotides downstream of the stop codon and this probe did not detect the high molecular weight band, consistent with the model that the PR mRNA ends some 10000 nucleotides downstream of the stop codon (**Fig 35d**).

Next X. Yue used qRT-PCR to measure the relative abundance of RNA throughout the PR coding locus and beyond. She measured relatively consistent levels of RNA through the annotated 8 exons of the PR mRNA and throughout the putative 3' UTR until around 10000 nucleotides downstream of the stop codon, where a sharp decrease of RNA transcript levels was observed. Again this result is consistent with the model that the PR mRNA ends around 10000 nucleotides downstream of the stop codon (**Fig 35e, Fig 38**).

Finally, X. Yue used 3' RACE to determine the 3' terminus of the PR mRNA. Primers were designed to target some 8000 and 9000 nucleotides downstream of the PR stop codon. X. Yue detected polyadenylation sites at +13029, +13032, +13035, +13038, and +13048 with respect to the mRNA +1 transcription start site. At the level of genomic DNA, these polyadenylation sites lay some 100000 base pairs away from the transcription start site (**Fig 36**).

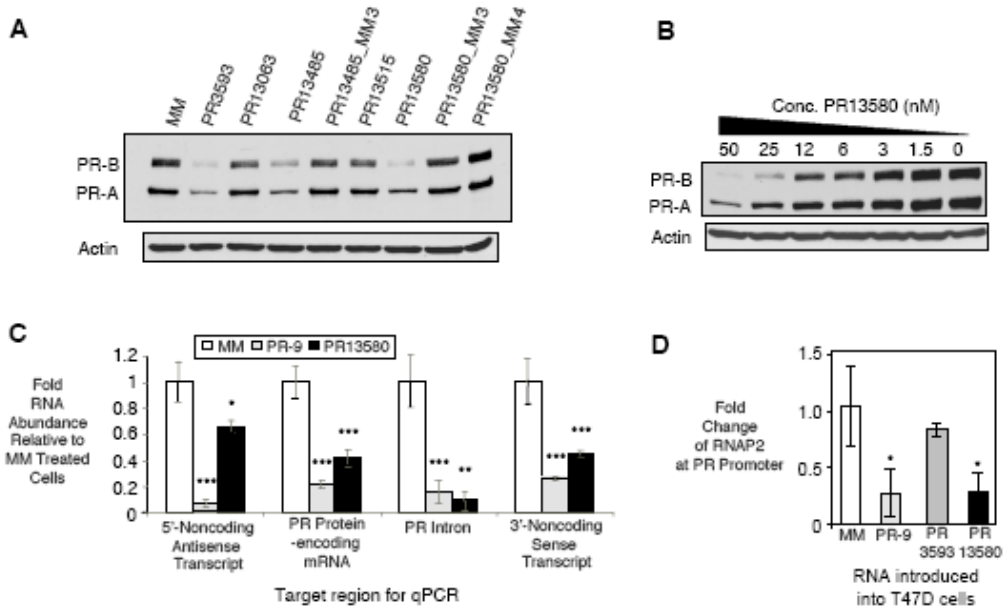


Figure 39: Inhibition of PR expression in T47D cells by agRNAs complementary to sequences downstream from the terminus of PR mRNA. (**A**) Western analysis showing inhibition of protein expression by duplex RNAs added to cells at a concentration of 50 nM. (**B**) Dose response for PR13580. (**C**) qPCR showing reduction of PR mRNA levels by duplex RNAs added at 25 nM. Four different primer sets were used, each complementary to different regions of the PR gene. (**D**) Presence of RNAP2 to the PR transcription start site (25 nM duplex RNA) evaluated by ChIP. ***p<0.005, **p<0.01, and *p<0.05 as compared to cells treated with a mismatch RNA. p-values were calculated using the two tailed unpaired Student's T-test with equal variances. All error bars represent standard deviation.

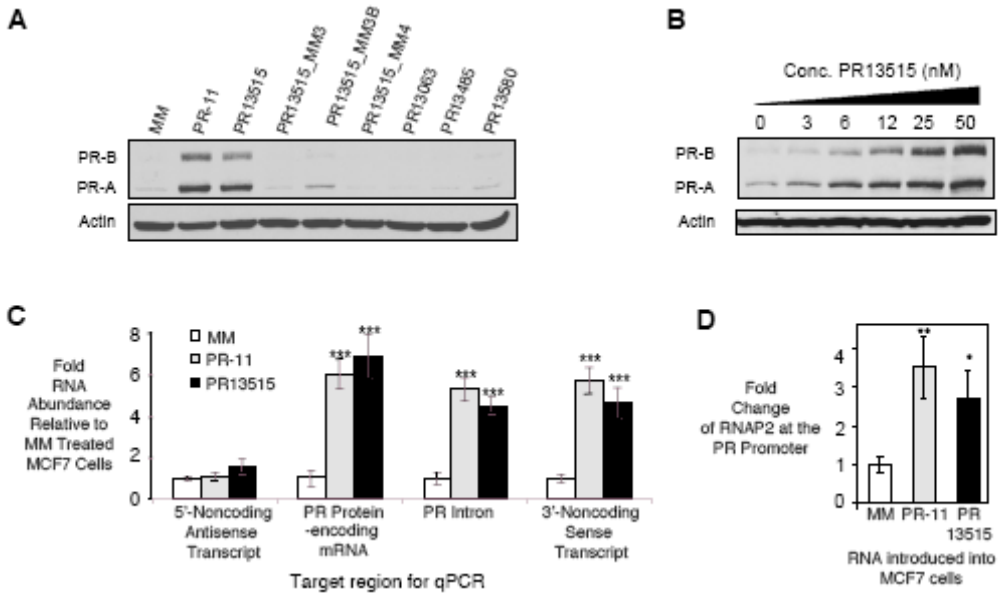


Figure 40: Enhanced PR expression in MCF7 cells by an RNA complementary to a sequence downstream from the 3'-UTR terminus in PR. (A) Western analysis showing activation of protein expression by duplex RNAs. (B) Dose response for RNA PR13515. (C) qPCR showing effect on RNA levels. Four different primer sets were used, each complementary to different regions of the PR gene. (D) Recruitment of RNAP2 to the PR promoter upon addition of PR13515 or PR-11 evaluated by ChIP.

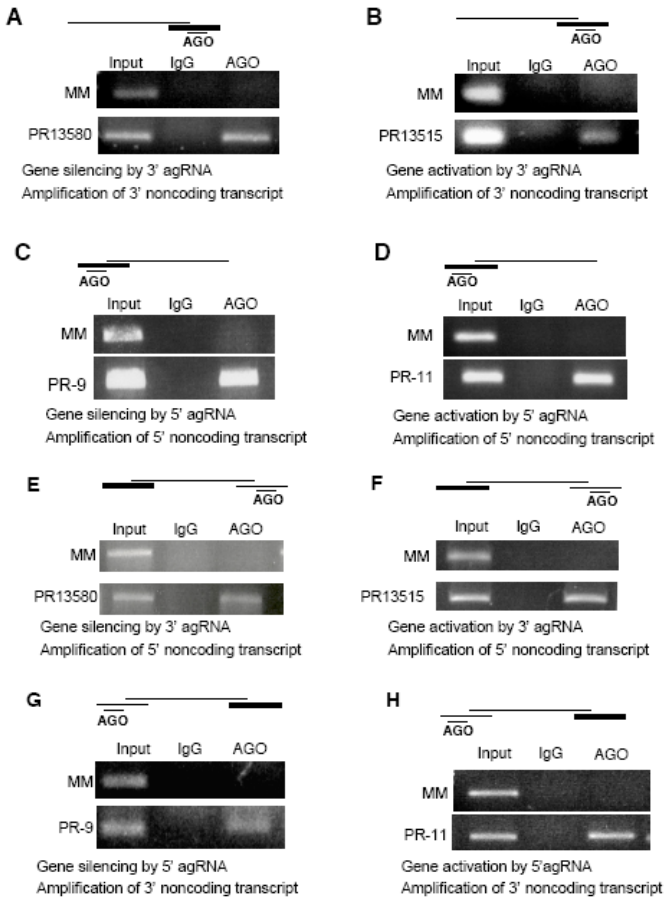


Figure 41: Effect of 3' or 5' agRNAs on recruitment of AGO protein to the 3' or 5' noncoding transcripts at the PR locus. Effect of adding (A) inhibitory RNA PR13580 to T47D cells or (B) activating RNA PR13515 to MCF7 cells on recruitment of AGO protein to the 3' noncoding transcript. Effect of adding (C) inhibitory RNA PR-9 to T47D cells or (D) activating RNA PR-11 to MCF7 cells on recruitment of AGO protein to the 5' noncoding transcript. Effect of adding (E) inhibitory RNA PR13580 to T47D cells or (F) activating RNA PR13515 to MCF7 cells on co-immunoprecipitaton of AGO protein with the 5' noncoding transcript. Effect of adding (G) inhibitory RNA PR-9 to T47D cells or (H) activating RNA PR-11 to MCF7 cells on co-immunoprecipitation of AGO protein with the 3' noncoding transcript. The scheme above each gel depicts PR mRNA, the 3' and/or 5' noncoding transcripts, and AGO bound agRNA. The heaviest line represents the transcript being amplified.

5.2 TRANSCIPTIONAL REGULATION BEYOND THE 3' TERMINUS OF PR

After identifying the 3' terminus of PR, X. Yue targeted duplex RNAs to regions beyond the 3' terminus of PR in MCF7 and T47D cells. She designed 4 duplexes targeting 15, 437, 467, and 532 nucleotides downstream of the most downstream PR polyadenylated site (**Fig 35f**). In order remain consistent between labeling of agRNAs targeting the PR promoter, it was decided to name these duplexes with respect to the mRNA transcription start site. Thus the duplexes targeting 15, 437, 467, and 532 nucleotides downstream of the PR 3' terminus were named PR13063, PR13485, PR13515, and PR13580. The 3' terminus of PR is again at PR13048. Again, at the level of genomic DNA, these duplex RNA target site were more than 100000 base pairs away from the PR promoter.

X. Yue transfected agRNAs PR13063, PR13485, PR13515, and PR13580 in T47D cells and observed potent reduction of PR gene expression for PR13485 and PR13580 (**Fig 39a**). Knockdown of PR expression by PR13580 is dose dependent and seen at the level of protein and mRNA (**Fig 39b, c**). To determine if PR silencing was at the level of transcription, I performed ChIP of RNAP2 (CTD4H8) with cells treated with either mismatched negative control, promoter targeting agRNA, siRNA (PR3593), or 3' targeting agRNA. The result showed that both promoter and 3' targeting agRNAs reduce RNAP2 occupancy at the PR transcription start site relative to negative control but siRNAs do not effect transcription (**Fig 39d**).

X. Yue also transfected agRNAs PR13063, PR13485, PR13515, and PR13580 in MCF7 cells and observed activation of PR expression for PR13515. Multiple different designs of mismatches of PR13515 were tested, some with mismatches clustered at the 5' or 3' ends of the duplex, assured that gene activation was a sequence specific phenomenon (**Fig 40a**). Activation was observed at the level of protein and mRNA (**Fig 40b, c**). I used RNAP2 (CTD4H8) ChIP to test if activation occurred at the level of transcription. Both 5' and 3' targeting agRNAs induced increased recruitment of RNAP2 to the PR transcription start site relative to negative mismatch control (**Fig 40d**).

5.3 INVOLVEMENT OF NONCODING RNAS IN 3' TARGETING agRNAS

X. Yue tested for the existence of an RNA to serve as the molecular target for 3' targeting agRNAs. Using 3' and 5' RACE, X. Yue discovered an unspliced sense transcript overlapping the 3' terminus of PR (**Fig 37a**). RT-PCR using primers directed within the PR mRNA and within the noncoding transcript suggests that the sense transcript is a unique transcript and not an extension of the PR mRNA **(Fig 37b**).

I used RNA immunoprecipitation to test if 3' agRNAs recruited argonaute proteins to interact with the sense noncoding transcript. I used an antibody that recognizes all four of the human argonaute proteins (2A8 (*288*)). I found that in T47D cells, the silencing agRNA PR13580 recruits argonaute proteins to the 3' sense transcript (**Fig 41a**). For a positive control, I used the association of 5' targeting agRNA, PR-9, with the antisense transcript AT-2 (**Fig 41c**). I performed RNA immunoprecipitation on MCF7 cells and found that the activating agRNA PR13515 recruits argonaute proteins (2A8) to the 3' sense transcript (**Fig 41b**). For a positive control, I used the association of 5' targeting activating agRNA, PR-11, with the antisense transcript AT-2 (**Fig 41d**). Y. Chu repeated these experiments with an antibody specific to Ago2 (clone 9E8.2, Millipore 04-084). There also he found that both activating and silencing agRNAs targeting the 5' and 3' ends of PR recruits Ago2 to their respective targets.

X. Yue used 5' RACE to try to detect a cleaved product of the 3' noncoding RNA after agRNA treatment. Cleaved PR mRNA was detectable after treatment with siRNA but no cleavage of the 3' noncoding RNA was observed for agRNAs (**Fig 42**).

5.4 LONG DISTANCE CHROMATIN INTERACTIONS AND PR

As mentioned earlier, the 3' end of the PR gene lies more than 100 kb away from the transcription start site for PR. Thus when argonaute binds at the 3' end of the PR gene, there must be a mechanism for regulatory signals to reach the transcription start site where RNAP2 recruitment is either upregulated or downregulated. I proposed that this long distance signaling may be mediated by long distance chromatin interactions.

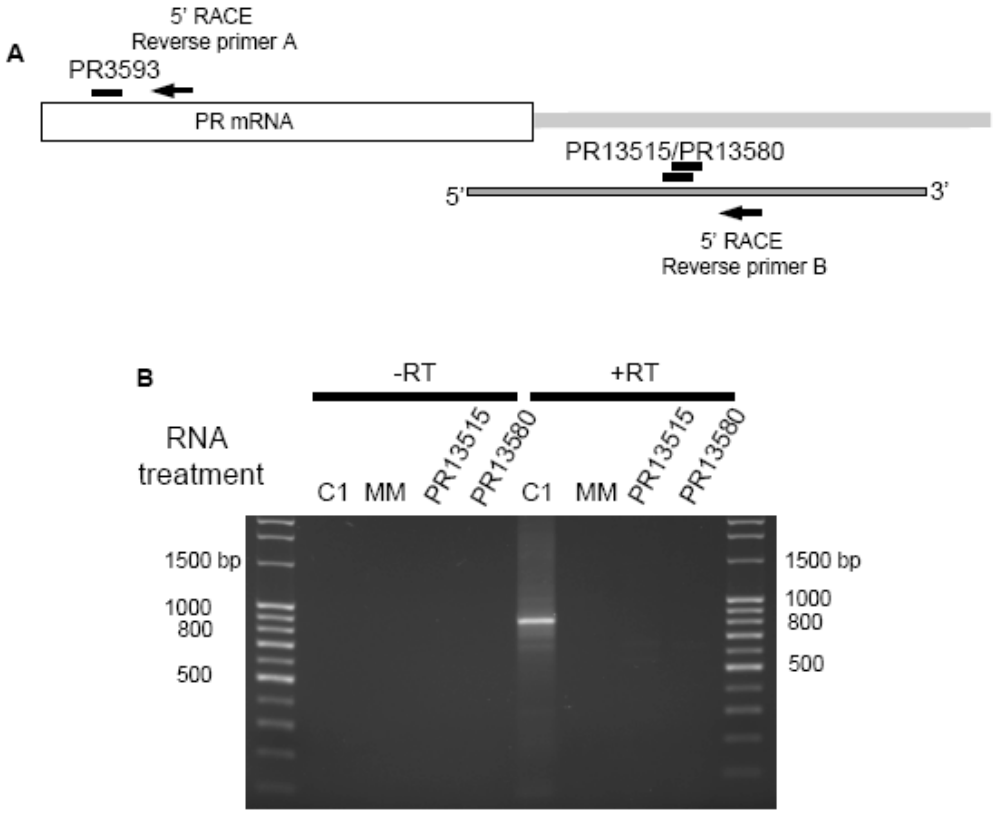


Figure 42: 5'-RACE showing no detection of cleavage products within the 3' noncoding transcript after addition of activating agRNA PR13515 or inhibitory agRNA PR13580. (A) Location of primers for 3' and 5' RACE and location of duplex RNAs PR3593, PR13515, and PR13580. (B) Results of 5'-RACE with primer B (5'-TCAACTCAAACTTACAGCAAGAATCCTGTTCCACTC-3') downstream from the recognition sites of PR13515 and PR13580 on 3' noncoding sense transcript didn't detect any cleavage site. As a positive control, the cleavage site by PR3593, an siRNA targeting PR mRNA, was detected with the downstream primer A (5'-AGAAACGCTGTGAGCTCGACACAACTCC-3'). The sense strand sequence of C1 is 5'-ATGGAAGGGCAGCACAACT-3'. (C) (next page) The sequence of PCR product from positive control sample treated with siRNA PR3593 was aligned to PR mRNA sequence by BLAST. The part of PR3593 sequence contained in the PCR product is highlighted.

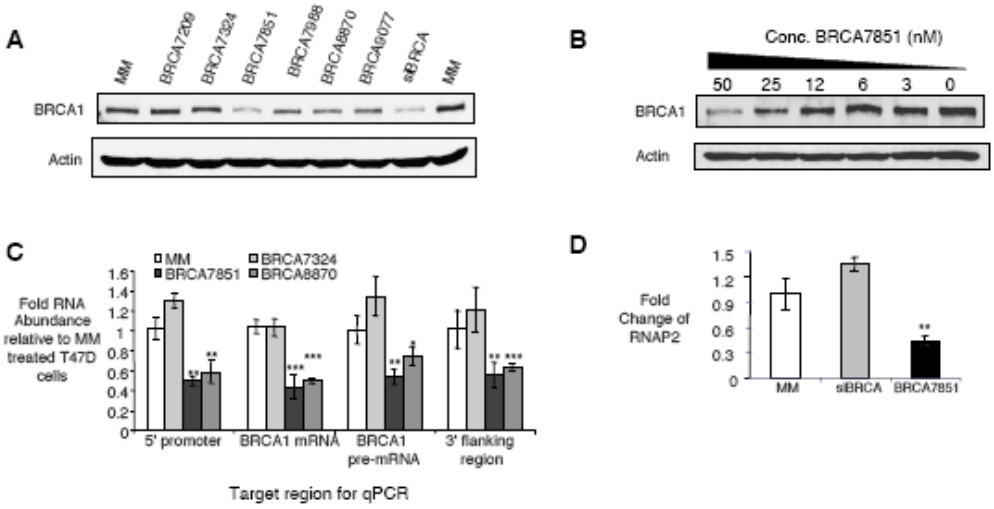


Figure 43: Inhibition of BRCA1 expression by agRNAs targeting sequences beyond the 3'-UTR. Western analysis showing levels of BRCA1 protein after addition of (A) six agRNAs complementary to sequences beyond the 3' polyadenylation site for BRCA1 mRNA and siBRCA1 complementary to BRCA1 mRNA. (B) increasing concentrations of BRCA7851. (C) qPCR analysis of RNA levels using primer sets designed to detect noncoding RNA at the BRCA1 promoter, BRCA1 mRNA, BRCA1 pre-mRNA, and noncoding RNA beyond the 3' terminus of BRCA1 mRNA. (D) Presence of RNAP2 at the BRCA1 transcription start site evaluated by ChIP. All error bars are standard deviation. ***$p<0.005$, **$p<0.01$, and *$p<0.05$ as compared to cells treated with RNA MM. p-values were calculated using the two tailed unpaired Student's T-test with equal variances. All error bars represent standard deviation. Duplex RNAs were added to cells at 25 nM unless otherwise noted.

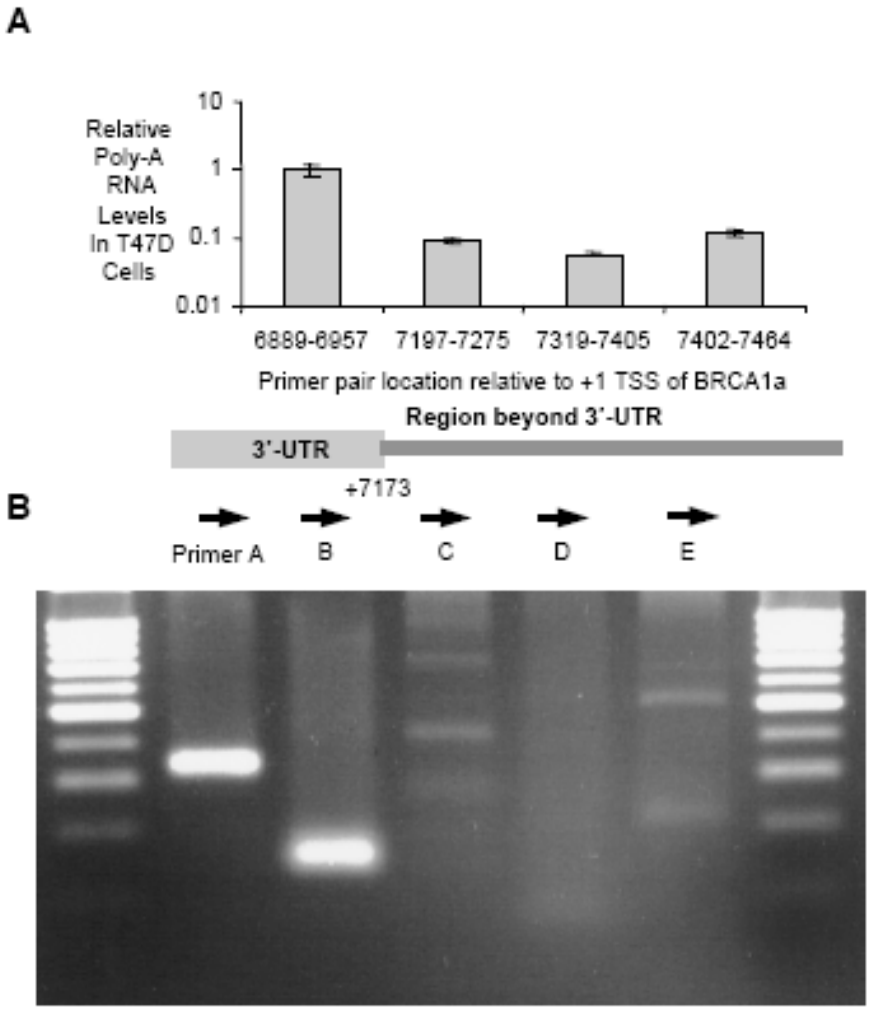


Figure 44: Measurement of poly-A RNA levels in T47D cells using different primer sets surrounding the 3' terminus of BRCA1 mRNA. (A) qPCR showing levels of poly-A RNA detected from the BRCA1transcription start site past the most downstream annotated terminus of PR mRNA. (B) 3' RACE using primers upstream and downstream of the termination site for BRCA1 revealed no evidence for a longer BRCA1 messenger RNA transcript. Primers A and B detected the known BRCA1 termination and primers C, D, and E downstream of that site did not detect amplified product.

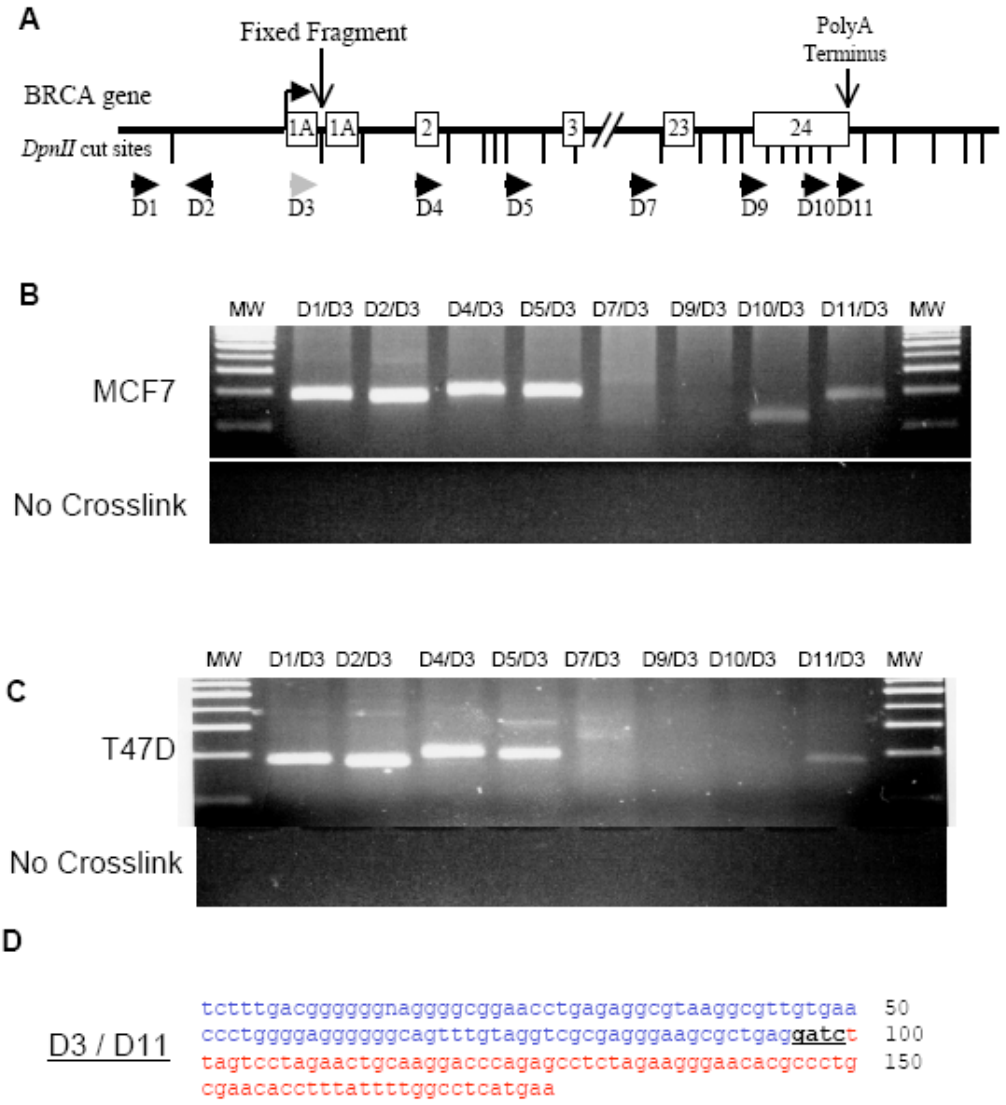


Figure 45: 3C analysis of the BRCA1 gene. (A) Primer sets used as described by Brown and coworkers (Tan-Wong et al., 2008). Primer D3 was a fixed reverse primer. (B) Agarose gel analysis of 3C products in MCF7 cells. In contrast to previous results (Tan-Wong et al., 2008) we oberved product when using primer D4 but not with primer D9. (C) 3C analysis for T47 D cells. (D) 3C product D11/D3 was cloned and sequenced. The product aligns with sequences at the 5' and 3' termini of the BRCA1 gene. Control experiments using non-crosslinked samples produced no PCR product. No Crosslink: no crosslinker added to sample prior to ligation and amplification.

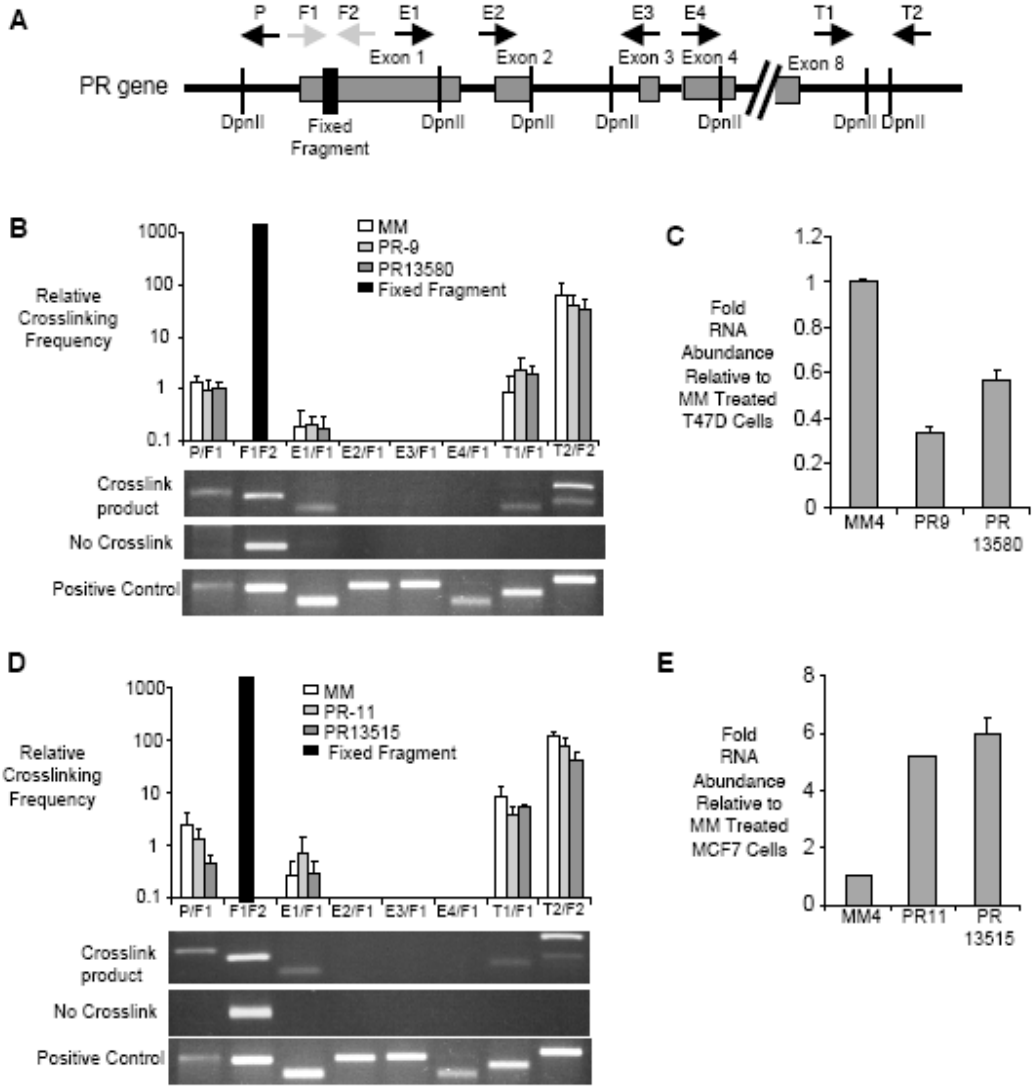

A
P
F1
F2
E1
E2
E3
E4
T1
T2
Exon 1
Exon 2
Exon 3
Exon 4
Exon 8
PR gene
DpnII
Fixed Fragment
B
MM
PR-9
PR13580
Fixed Fragment
Relative Crosslinking Frequency
1000
100
10
1
0.1
P/F1
F1F2
E1/F1
E2/F1
E3/F1
E4/F1
T1/F1
T2/F2
Crosslink product
No Crosslink
Positive Control
C
Fold RNA Abundance Relative to MM Treated T47D Cells
1.2
1
0.8
0.6
0.4
0.2
0
MM4
PR9
PR 13580
D
MM
PR-11
PR13515
Fixed Fragment
Relative Crosslinking Frequency
Crosslink product
No Crosslink
Positive Control
E
Fold RNA Abundance Relative to MM Treated MCF7 Cells
8
6
4
2
0
MM4
PR11
PR 13515

Figure 46: 3C analysis of the PR locus. (A) Schematic of the PR gene showing DpnII cleavage sites, exon boundaries, and locations of primers used for 3C analysis. The primer pairs used for 3C amplification are shown on the x axes of parts (B) and (D). (B) Top, qPCR showing the relative levels of detection of crosslinked product after treatment with a mismatch-containing RNA duplex or inhibitory duplexes PR-9 or PR13580. (C) qPCR showing reduced RNA levels in samples used for part (B). (D) Top, qPCR showing the relative levels of crosslinked product after treatment with a mismatch-containing RNA duplex or activating duplexes PR-11 or PR13515. (E) qPCR showing increased RNA levels in the samples used for part (D). Primer P amplifies a sequence at the PR promoter. Primers E1, E2, E3, and E4 amplify sequences within PR exons 1-4. Primers T1 and T2 amplify sequences beyond the terminus of PR mRNA. F1/F2= Fixed fragment. As described by Baylin and coworkers (Tiwari et al., 2008), the fixed fragment is a normalization control derived from genomic DNA by primers complementary to sequences with exon 1. The bar represents performance of the normalization control, not its absolute value. Values in parts (B) and (D) are relative to amplification of sequence at the PR promoter using primer P. Duplex RNAs were added to cells at 25 nM. Two bands are observed in the T2/F2 analysis because of an alternative DpnII cleavage site. The positive control shows amplification of an analogous single stranded DNA.

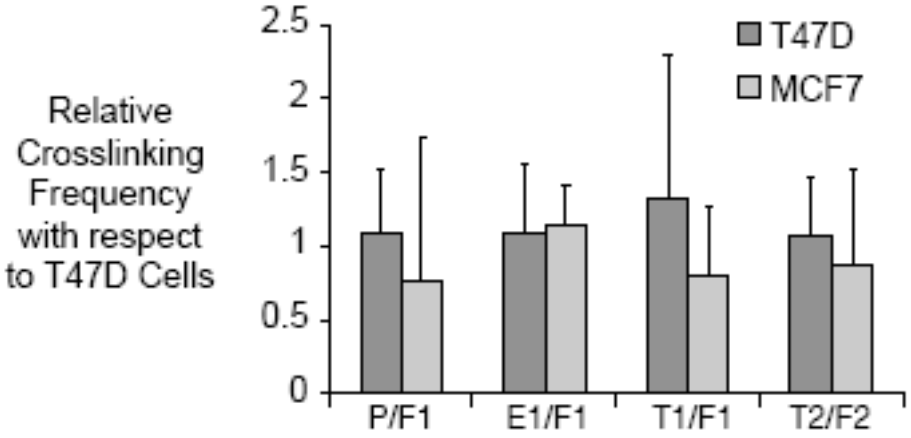


Figure 47: Comparing looping in MCF7 versus T47D cells. Crosslinking frequencies are relative to detection of a fixed fragment within genomic DNA for untreated T47D and MCF7 cells. Primer set Ex1 amplify PR exon 1-4. Primer sets T1 and T2 amplify sequences beyond the terminus of PR mRNA.

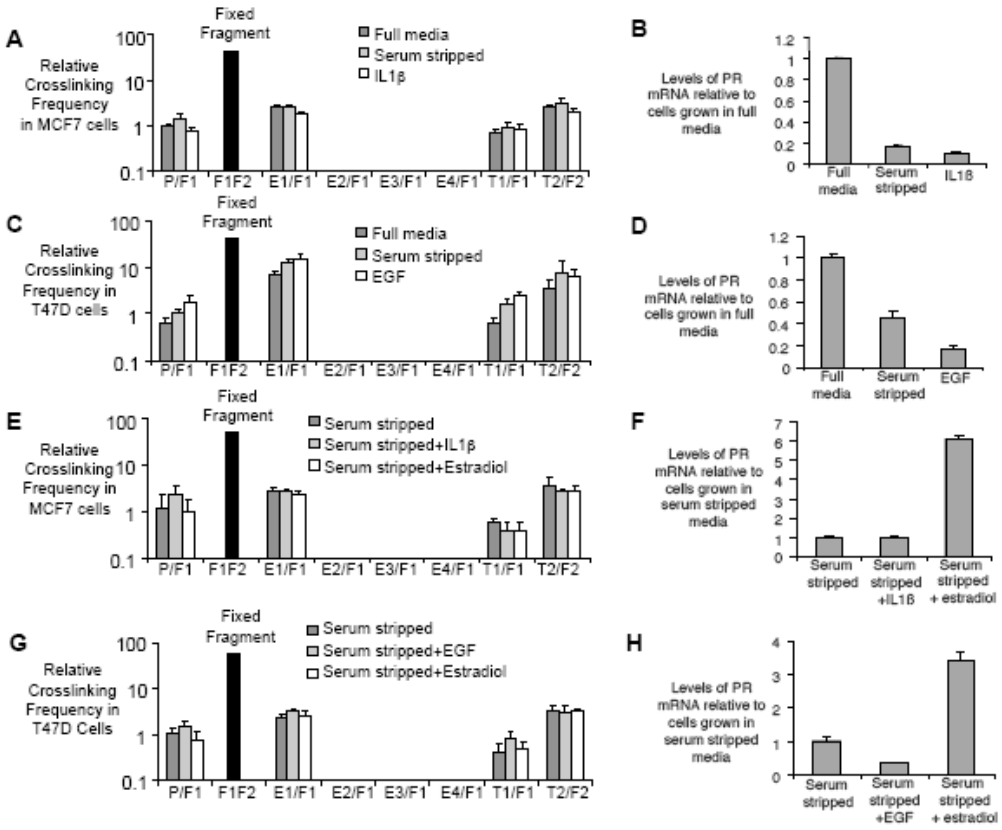


Figure 48: 3C analysis of MCF7 and T47D cells using primers sets as described in Figure 7A. 3C analysis for MCF7 (A) or T47D (C) cells upon treatment with full media, serum-stripped media, IL1ß (MCF7), or epithelial growth factor (EGF, T47D). 3C analysis for MCF7 (E) or T47D (G) cells upon treatment with serum-stripped media or serum-stripped media supplemented with IL1ß (MCF7), or epithelial growth factor (EGF, T47D), or estradiol (both T47D or MCF7). Parts (B), (D), (F), and (H) show the effects of treatments on PR mRNA levels as monitored by qPCR. FF= Fixed Fragment. As described by Baylin and coworkers,the "Fixed Fragment" is a normalization control derived from genomic DNA by primers complementary to sequences with exon 1. The bar represents performance of the normalization control,Anot its absolute value. Primer combinations are those shown in Figure 7A.

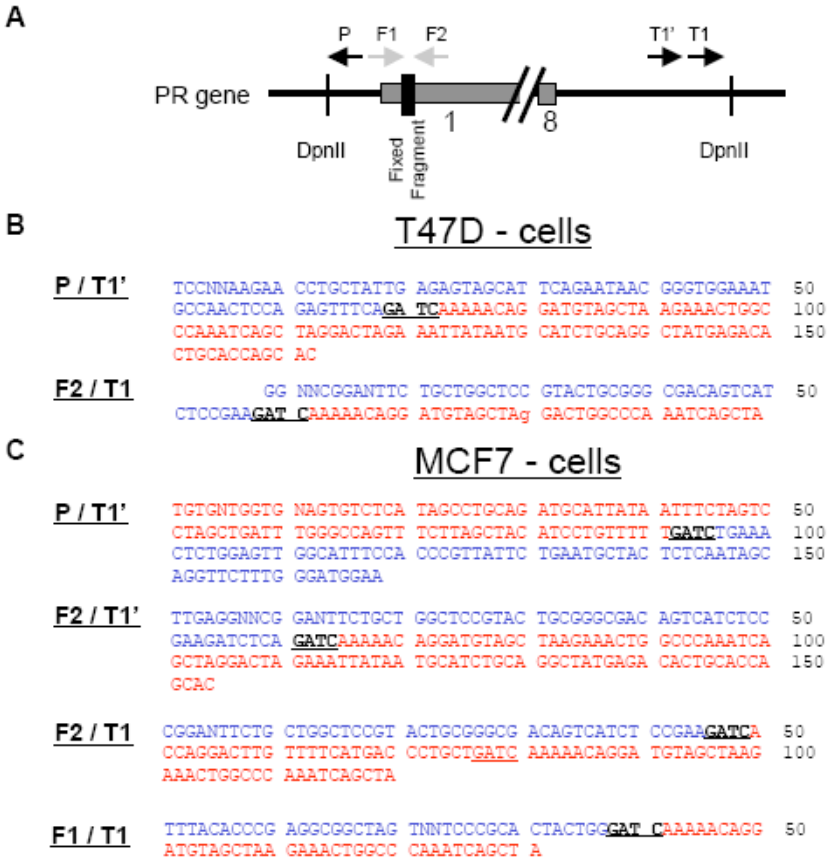


Figure 49: PCR product was sequenced for some combinations of primers detecting gene loops between the 5' and 3' ends of PR. (A) Product was obtained four possible combinations primers between two cut sites at the PR 5' end and one cut site and the PR 3' end. (B) Two PCR products were obtained in T47D cells and correctly aligned with their genomic targets at the 5' and 3' ends of PR. Blue text aligned with the 5' end and red text aligned with the 3' end. The DpnII cut site used is bold and underlined. (C) Four PCR products were obtained in MCF7 cells and correctly aligned with their genomic targets at the 5' and 3' ends of PR. Coloring is as in part B. Note for FF2 / T1 the use of a nearby DpnII cut site for ligation. The alternative site is underlined but not bold. The use of alternative of cut sites explains additional bands seen by gel electrophoresis of PCR product (see Figure 7B and 7D). Alignments were made using BLAT from UCSC genome brower at genome.ucsc.edu.

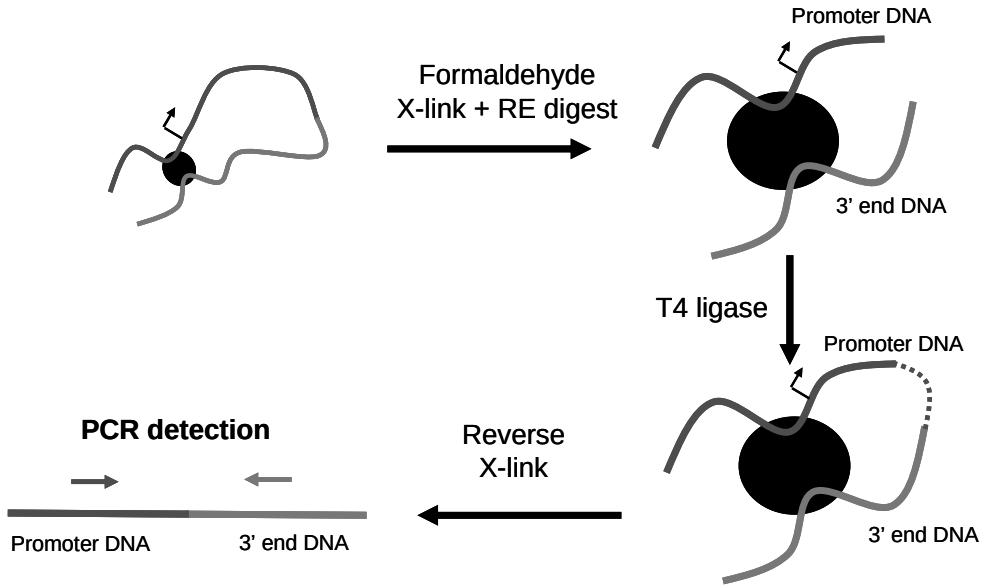


Figure 50: Scheme for 3C experiment. Chromatin conformation capture (3C) is a crosslinking based assay where DNA fragments held together by protein interactions are trimmed and ligated together. The resulting ligation product is detected by PCR.

To help develop this hypothesis, X. Yue tested 3' targeting agRNAs targeting the gene BRCA1. This gene had already been shown to have a long distance interaction between the 5' and 3' ends of the gene (*294*). Using 3' RACE, I confirmed that the 3' end of BRCA1 was as previously reported (**Fig 44b**). X. Yue tested agRNAs targeting 36, 151, 678, 815, 1697, and 1904 nucleotides downstream of the BRCA1 polyadenylation site. The BRCA1 mRNA is 7173 nucleotides long, thus with respect to the BRCA1 transcription start site, these duplexes were named BRCA7209, BRCA7324, BRCA7851, BRCA7988, BRCA8870, and BRCA9077. These duplexes were transfected into T47D cells. The duplex BRCA7851 caused marked silencing of BRCA1 expression and duplexes BRCA7988, BRCA8870, and BRCA9077 caused slight reduction of BRCA1 protein (**Fig 43a**). Silencing by duplexes BRCA7851 and BRCA8870 was seen at the level of protein and mRNA (**Fig 43c**). Silencing of BRCA7851 was also dose dependent (**Fig 43b**). X. Yue performed a ChIP of RNAP2 and found that unlike siRNAs, BRCA7851 silences at the level of transcription (**Fig 43d**).

I used a technique known as chromatin conformation capture (3C) to detect long distance chromatin interactions (**Fig 50**). The premise of this technique is that DNA held in close proximity by protein:protein interactions can be formaldehyde crosslinked together. After formaldehyde crosslinking, DNA not held in complex is trimmed away using restriction enzymes. Lastly, DNA is treated with T4 DNA ligase in a very dilute solution to allow DNA ends held together by formaldehyde crosslinks to be ligated together. This creates a unique PCR product that can only be produced by the presence of protein interactions hold the distant DNA pieces together.

First I verified the previously reported existence of chromatin interactions between the 5' and 3' ends of the BRCA1 gene in MCF7 and T47D cells (*294*). I formaldehyde crosslinked untreated cells and purified nuclei. I digested chromatin with the restriction enzyme DpnII and used the same primer sequences reported by Tan-Wong et al (**Fig 45a**). As Tan-Wong reported in both cell lines I detected interactions between the 3' terminus and the first exon of the BRCA1 gene (primer set D11/D3). In contrast to Tan-Wong's report, I detected chromatin interactions between exon 2 and exon 1 also (primer set D4/D3). I did not see evidence of interactions between intron 23 and exon 1

(primer set D9/D3) and I only detected interactions between exon 24 and exon 1 in MCF7 cells (primer set D10/D3) (**Fig 45b, c**).

For a negative control I performed 3C on cells not crosslinked with formaldehyde to test for nonspecific detection of distant chromatin interactions. Cells not treated with formaldehyde did not detect any DNA interactions. Also, X. Yue cloned and sequenced the product from primer set D11/D3 detecting the interaction and this product did correspond to the correct sequences at the 3' and 5' ends of the BRCA1 gene ligated at a DpnII restriction site (**Fig 45d**).

I then performed 3C to detect interactions between the 5' and 3' ends of the PR gene. The nearest DpnII cut site to the 3' targeting agRNA site is located approximately 500 nucleotides away (detected by primers labeled T1). The nearest cut site to the PR transcription start site is approximately 150 nucleotides away from the TSS (detected by primers labeled F1 and F2) (**Fig 46a**). 3C detected interactions between the PR TSS and the promoter, exon 1, and two 3' terminal sites near the 3' agRNA target site (either PR13515 or PR13580) in both T47D and MCF7 cells (**Fig 46b, d**). For a negative control I tested cells not treated with formaldehyde and found no detection of DNA interactions. For a positive control, I ordered the correct PCR amplicon produced by ligation of DpnII cut sites synthesized (Sigma-Aldrich). I performed PCR on this synthesized product and found that the primer could amplify the PCR product.

I sequenced products of primers targeting the sites T1, the PR transcription start site, and the PR promoter. In both T47D and MCF7 cells these products corresponded to the correct DNA sequences ligated together at a DpnII cut site (**Fig 49**). I found that digestion by the restriction enzyme DpnII is not 100% efficient, likely inhibited by formaldehyde crosslinking. In one sequenced product, a nearby DpnII cut site was preferred resulting in the inclusion of two DpnII cut sites in the PCR amplicon.

I performed qRT-PCR of 3C product to test if the frequency of 5' and 3' interactions changed upon treatment with either 5' or 3' targeting agRNAs. Using T47D cells treated with silencing agRNAs PR-9 or PR13580 revealed no significant change in the frequency of DNA interactions (**Fig 46b)**. Similarly, in MCF7 cells treated with activating agRNAs PR-11 or PR13515 revealed no significant change in DNA

interactions (**Fig 46d**). Additionally, the relative abundance of 5' to 3' loops in MCF7 and T47D cells was not different for the PR gene dispite their significantly different levels of transcription (**Fig 47**).

In comparison, some reports have suggested that stimulation by hormone treatment may modulate long distance DNA interactions. I collaborated with B. Janowski to test if this occurred in the case of PR. B. Janowski treated T47D cells with either regular media (full media), media where the serum had hydrophobic molecules removed by charcoal filtering (serum-stripped), epithelial growth factor (EGF), serum-stripped media and EGF, or serum-stripped media with estradiol for 24 hours before treating with formaldehyde and purifying nuclei. B. Janowski treated MCF7 cells with either full media, serum-stripped media, IL1β, serum-stripped media with IL1β, or serum-stripped media with estradiol before treating with formaldehyde and purifying nuclei.

These stimulations produced the expected changes in PR expression (**Fig 48b,d,f,h**). Treatment of cells with serum-stripped media, EGF, and IL1β all reduced PR expression. Estradiol treatment increases PR expression. I performed 3C experiments in cells treated by all of these stimuli and saw no change in DNA interactions for either MCF7 or T47D cells (**Fig 48a,c,e,g**). This result agrees well with the result that DNA interactions are not altered by gene silencing or activation through agRNAs either. Instead, DNA remains in a fixed loop which may allow protein interactions at the 3' end of the PR gene to signal changes in RNAP2 recruitment at the PR transcription start site.

I tested for differences in chromatin modifications with 3' and 5' targeting agRNAs. I used ChIP with an antibody targeting H3K27 trimethylation (Millipore 07-449) in cells treated with either MM or silencing or activating agRNAs. In T47D cells, 3' and 5' silencing agRNAs induced increased H3K27me3 marks (**Fig 51a**). In MCF7 cells, 3' and 5' activating agRNAs induced decreases in H3K27me3 marks (**Fig 51b**). H3K27 trimethylation is a marker that is associated with polycomb protein recruitment. The mammalian polycomb group proteins CBX7 and CBX8 contain chromobox domains similar to *S. pombe* ChP1, the member of the RITS complex, but otherwise have no other homology (*295-297*).

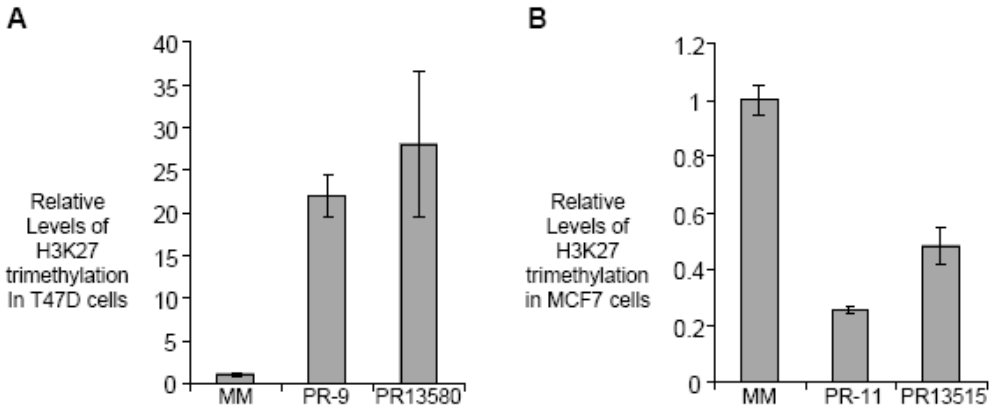


Figure 51: ChIP of H3K27me3. Chromatin immunoprecipitation for H3K27 trimethylation (Millipore) marker within the PR gene locus. (A) Gene silencing in T47D cells is associated with increased in levels of H3K27 trimethylation marker for agRNAs PR-9 and PR13580. (B) Gene activation is associated with decrease in H3K27 trimethylation marker for agRNAs PR-11 and PR13515.

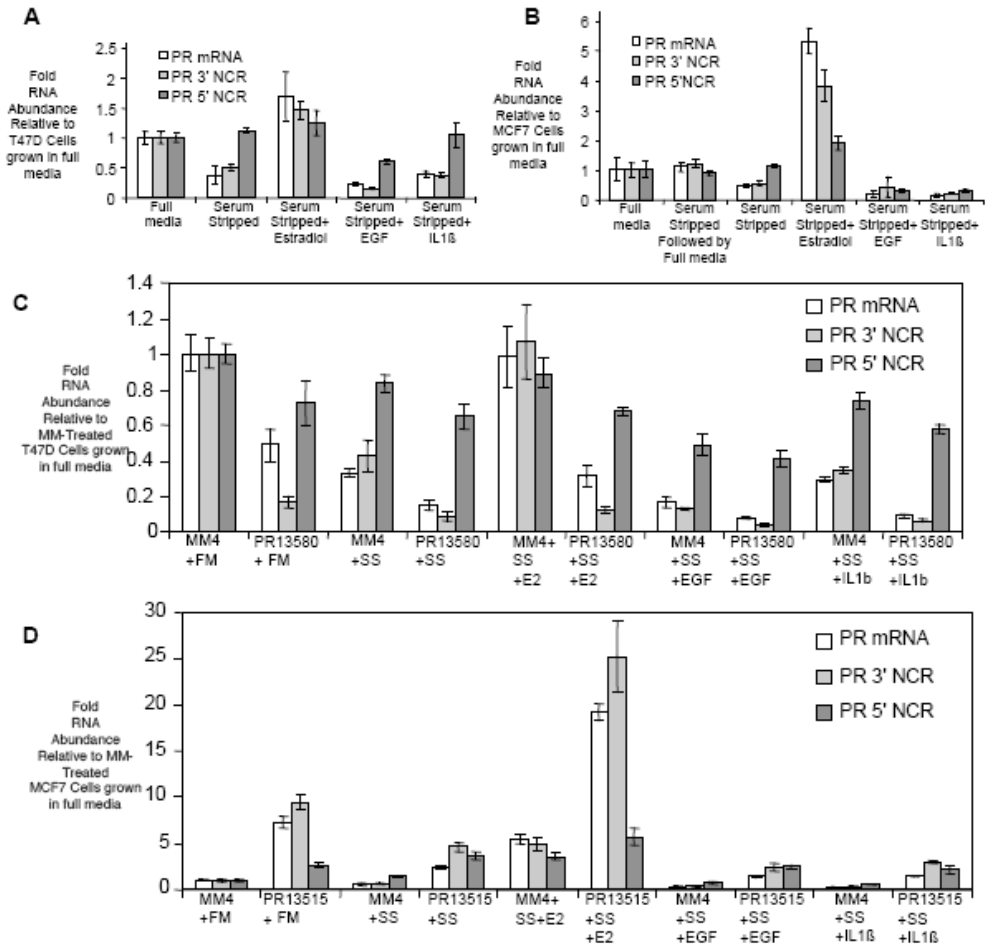


Figure 52: Effect of physiologic stimuli or effect of combining physiologic stimuli with addition of agRNAs on expression of PR mRNA, the 3' noncoding RNA, and the 5' noncoding RNA. qPCR analysis of the effect of physiologic stimuli on transcript expression in (A) T47D cells and (B) MCF7 cells. qPCR analysis of the effect of physiologic stimuli and agRNA addition on transcript expression in (C) T47D cells and (D) MCF7 cells. PR 3'NCR: 3' noncoding PR RNA. PR 5'NCR: 5' noncoding PR RNA. Serum stripped followed by full media: Cells were grown in serum stripped media, the media was replaced by full media, and subsequently harvested. E2: 17ß estradiol treatment. IL1ß: interleukin 1ß treatment. EGF: epidermal growth factor treatment.

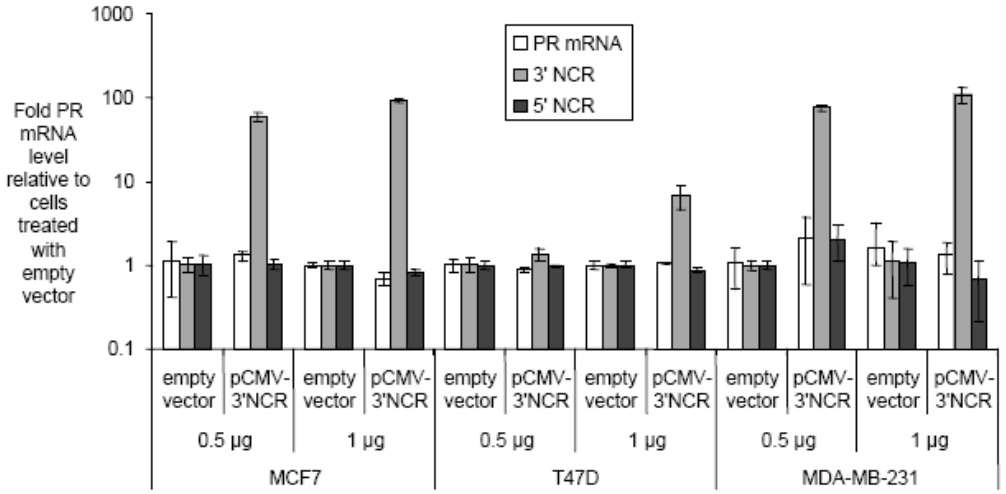


Figure 53: Effect of overexpression of the 3' noncoding transcript on expression of transcripts at the PR locus. qPCR analysis of the effect of overexpression of the 3' noncoding transcript on PR mRNA, the 5' antisense noncoding transcript, and the endogenously expressed 3' noncoding transcript.

5.5 NONCODING RNAS AT THE 5' AND 3' ENDS OF PR INTERACT

If chromatin is held in a loop conformation bringing together the 5' and 3' ends of the PR gene, it is possible that noncoding RNAs associated with chromatin are thereby associated with together. I tested the hypothesis that an RNA immunoprecipitation of an argonaute interaction (2A8) with the noncoding RNA at one end of the PR gene will also pulldown the noncoding RNA at the other end of the PR gene.

For agRNAs targeting the 3' end of PR, activating RNA PR13515 in MCF7 (**Fig 41f**) or silencing RNA PR13580 in T47D (**Fig 41e**), I found that they also pull down the noncoding RNA AT-2 at the 5' end of the gene. Similarly, for agRNAs targeting the PR promoter, activating agRNA PR-11 in MCF7 (**Fig 41h**) or silencing agRNA PR-9 in T47D (**Fig 41 g**), I found that they pull down the noncoding sense transcript at the 3' end of the PR gene. This suggests that these noncoding RNAs are part of a protein or chromatin based complex linking the 5' and 3' ends of the PR gene and thereby an argonaute interaction with one noncoding RNA is an argonaute interaction with both. Y. Chu also repeated these results with an Ago2 specific antibody (clone 9E8.2)

5.6 PHYSIOLOGICAL RELEVANCE OF THE 3' SENSE TRANSCRIPT

K. Gagnon and B. Janowski collaborated to study the how the 3' and 5' noncoding transcripts respond to physiological stimuli. They treated T47D cells with full media, media stripped of hydrophobic molecules, IL1β, EGF, and estradiol. IL1b, EGF, and serum stripped media induced silencing of both PR and the 3' noncoding RNA. Estradiol activated both PR and the 3' noncoding RNA. The 5' noncoding RNA response to these stimuli was relatively diminished compared to the PR response (**Fig 52a-d**).

K. Gagnon then constructed an expression plasmid to express the 3' sense transcript from a CMV promoter. B. Janowski transfected these plasmids into T47D cells, MCF7 cells, and MDA-231 cells, which do not express PR. In all cases, overexpression of the 3' sense transcript in *trans* had no effect on PR expression or the 5' antisense transcript (**Fig 53**).

5.7 GENE LOOPS AS A GENERAL MODEL FOR GENE TRANSCRIPTION

A small but growing body of literature suggests that gene looping between the 5' and 3' ends of genes might be a general feature of transcription (*298*). Initially, several yeast genes were shown by 3C to interact between their promoter and 3' termini (*299, 300*). The Ansari et al report is one of the earliest examples of the 3C protocol reported in literature and is unfortunately short on controls. The only positive control given is standard PCR of genomic DNA as a sort of loading control. Gene looping is shown to be higher in conditions in which gene transcription is activated for genes responsive to temperature and dextrose. However, this may be due to global changes in chromatin rather than a transcription regulatory mechanism.

O'Sullivan et al present 3C evidence using two different restriction enzymes that gene looping occurs for two genes in yeast. They go on to argue that the enhanced association of RNAP2 with the gene promoter and the gene termini compared to middle regions of the gene is evidence of gene looping (*299*). However, this enhanced association of RNAP2 at the gene terminus can also be explained by RNAP2 slowing as it processes the 3' end of mRNA and adds the poly(A) tail. Similarly, Proudfoot argues (*298*) that association of transcription factors TFIIB and TFIID with polyadenylation factor CPSF and SSU72 is evidence of gene looping (*301, 302*). This model may work for simple yeast genes. For mammalian genes it is likely overly simplified since it is difficult to explain how multiple transcription start sites and multiple polyadenylation sites are tolerated from the same gene locus, especially for large genes like PR and BRCA1 where dozens of transcription start sites and polyadenylation sites are spread over 100000 bases. Another explanation would be that these polyadenylation factors associate with elongating RNAP2 very early in transcription while RNAP2 is still near the promoter and sit poised on the CTD to terminate transcription.

In a recent study, Singh et al show gene looping by 3C for two genes in yeast (*303*). This 3C experiment was not very well controlled. An association of the transcription factor TFIIB with the 3' terminus of a gene in yeast is taken as further

evidence of gene looping. Here again, the story is likely more complicated than that since the pre-initiation complex factor TBP interacts strongly with the promoter but has no interaction with the 3' terminus. Finally, for the gene HIV-1 integrated in the human genome at two different loci, Perkins et al show strong evidence of gene looping between the promoter and termini by 3C using two different restriction enzymes (*304*). ChIP of RNAP2 at sites across the HIV-1 shows strong peaks of RNAP2 association at the 5' and 3' ends of the gene. However, here again, there is no evidence that the same RNAP2 molecule is associated with both ends. Instead, this may only be evidence of RNAP2 pausing at the 3' terminus.

Tan-Wong et al report that gene loops are affected by transcription activation. Unlike the Proudfoot model, Tan-Wong report that activating BRCA1 expression with estrodiol treatment causes gene loops to disassociate. Tan-Wong et al perform 3C using three different restriction enzymes. In mouse, pregnant mammary glands express high levels of BRCA1 and lactating mammary glands express low levels of BRCA1. In this case, lactating mammary glands are found to form gene loops but pregnant mouse mammary glands do not form gene loops. While 3C experiments seem to indicate that BRCA1 does form gene loops, the activation data is problematic. In our MCF7 and T47D cell lines, BRCA1 is very highly expressed without estradiol expression, so it is difficult to see how further activation with estradiol should induce dramatic chromatin reorganization. Furthermore, Tan-Wong et al only show activation by estradiol of a luciferase gene driven by a BRCA1 promoter and not of endogenous BRCA1 itself.

A recent thorough study by Tiwari et al concluded that long distance chromatin interactions in human cell lines remained fairly consistent as gene expression is turn on or off (*305*). However, between the states of gene transcription being turned completely off or on varied the tightness of packing. For the gene GATA-4, in a cell line where GATA-4 is turned off, the same pattern of chromatin interactions is seen but the relative frequency of detecting the interaction is stronger. This means that using semi-quantitative PCR, the bands detecting the chromatin interaction were brighter indicating that more T4 ligation occurred presumably due to the tighter packing (n=6 in each cell line). For this semi-quantitative PCR, standard curves were to performed to ensure quantization was linear. In

addition to tighter packing, chromatin markers, polycomb protein recruitment, and DNA methylation are all observed. These results agree well with our observation that chromatin loops change little as PR transcription is turned up or lowered.

Taken together, gene loops beween 3' termini and gene promoters remain an attractive model. It is likely that these genes were 3C is used, gene loops may actually form. It is helpful to think that 3' termination signals can be effectively conveyed to the gene promoter due to their close proximity. Furthermore, RNAP2 can cycle quickly from the 3' terminus back to the gene promoter when they are held close together, much like the "race track" model for ribosome translation of looped mRNAs. However, data concerning polyadenylation complexes, RNAP2 association, and transcription factors at both the 3' and 5' ends of the gene are difficult to interpret due to the complexity of transcription and especially since eukaryotic genes are characterized by many coding and noncoding transcription start sites and polyadenylation sites clustered at both the 5' and 3' ends. Data about gene loops being a general transcription regulatory mechanism remains murky. Hopefully future creative experiments will illuminate this story further.

5.8 COMPLEX TRANSCRIPTION FROM THE PR LOCUS

X. Yue carefully performed 3' and 5' RACE using primers targeting regions from exon 6 through the 3' terminus at +13000. In addition to the known PR-B and PR-A transcription start sites, she found 9 more sense transcription start sites situated in every PR exon except for exon 3 (**Fig 36**). These transcripts were spliced identically to the PR-B and PR-A mRNA and extend at least to a couple hundred nucleotides beyond the stop codon. Additionally, she found a large number of polyadenylation sites that clustered around 5 main sites in intron 7 and across exon 8. The PR coding region is rich in AUG start codons suggesting that any of these shorter PR transcripts still maintain a considerable amount of coding potential.

Real time PCR results vary from primer set to primer set but allow that full length PR may make up only half to a third of total accumulated spliced and polyadenylated transcripts from the PR locus (**Fig 35e, Fig 38**). Due to the size of the full

length PR mRNA at 13000 nucleotides, the transfer step for Northern protocols must be optimized to detect either longer (> 6000 nucleotides) or shorter transcripts (< 6000 nucleotides). This technical reason makes comparisons of relative abundance between longer and shorter RNA transcripts difficult by Northern. However, Northern experiments do detect a significant number of shorter PR transcripts whether probes to exon boundaries 2/3, 4/5, or 7/8 are used (**Fig 16c, Fig 35c,d**). These shorter RNA transcripts have also been reported in previous literature (*281, 306*).

Many splice variants and alternative isoforms of PR identical to those seen by X. Yue have been reported in the literature. There is no dispute that PR-B and PR-A are the major isoforms (*307*). Northern analysis identifies at least 6 unique PR RNA transcripts in T47D cells (*306*). Evidence even suggests that some of these downstream transcription start sites may even extend to include the long 3' UTR characterized by X. Yue (*308*). PCR analysis suggests the existence of even more variants and splice isoforms, several of which induce frame shifts which induce early C-terminal truncations of the PR coding region (*309, 310*).

The most prominent putative N-terminal truncated isoform of PR is named PR-C with a protein molecular weight of between 60 and 65 kd (*311-313*). Unfortunately, many of the studies of putative isoforms of PR depend heavily on forced overexpression of the putative isoform rather than studies of the endogenous protein's physiological importance (*309, 312*). This type of study suggests that the putative protein might have a physiological role if it were translated but gives no evidence that the actual protein exists. For PR-C, on the other hand, some evidence has been given that the endogenous protein may play a role in repressing PR-B transcriptional regulation during labor (*313*). That said, the existence of most of the putative isoforms of PR, including PR-C, remains hotly contested and it is not clear the exact amino acid sequence or corresponding mRNA for alternative PR proteins and mRNAs detected by western and northern analysis (*314*).

The difficulty with variants of PR that are detected by PCR is that the amplification step is powerful enough to detect RNA variants that may be transcriptional mistakes, which would never have a chance to be translated in the cytoplasm. Certainly, frameshifts induced by splice isoforms that prematurely truncate the coding potential are

prime targets for rapid degradation by the process of nonsense-mediated decay (*315, 316*). While the precise mechanism of this pathway is not fully understood, it is known that many mutated and incorrectly spliced transcripts and pseudo-genes in eukaryotes are allowed to persist in the cell when this RNA degradation pathway is removed. A model proposed is that these transcripts are allowed to be exported from the nucleus and after the first round of translation, the mistake is recognized, translation silenced, and the RNA is tagged for decay (*317*).

In addition to pathways that mark inappropriately processed mRNAs for decay, there is evidence that many polyadenylated RNAs are never allowed to be translated because they are not allowed to escape the nucleus (*318*). Some early studies suggested that the majority of polyadenylated transcripts meet their end in the nucleus and never escape to the cytoplasm (*319, 320*). While not understood or appreciated at the time, today sequencing and microarray studies have shown that nearly all of the eukaryotic genome is transcribed and many of these transcripts are sequestered to the nucleus (*203, 205, 207, 224*).

While being transcribed a nascent RNA transcript is coated with hnRNP proteins (*321, 322*). These proteins have roles in aiding in capping, splicing, and polyadenylation (*287, 298, 323*). After transcription, these proteins accompany the processed mRNA to the nuclear pore. There some pass through the pore with the mRNA but many are stripped away to remain in the nucleus as the mRNA is threaded through the nuclear pore and immediately bound by waiting ribosomes on the cytoplasmic side (*324, 325*). While some of these proteins that bind the nuclear RNA transcript play a role in making sure the RNA transcript is exported, other proteins bind an RNA transcript to ensure certain noncoding or mutated RNA transcripts will never be exported (*326-328*). These transcripts meet their end, sometimes very quickly, in the 3' to 5' exosome (*329, 330*).

Another observation by X. Yue about PR mRNA is heterogeneity in the site of polyadenylation clustered around a particular terminus site (**Fig 36**). For the long PR 3' UTR, there are at least 6 alternative polyadenylation sites clustered within 20 nucleotides of each other. For the short PR 3' UTR, there are some 16 polyadenylation sites clustered within 400 nucleotides of each other. How an elongating RNAP2 decides to use one

polyadenylation site or another to terminate a RNA transcript is not well understood. Polyadenylation is achieved when the RNA transcript is cleaved from the nascent RNA chain and an RNA polymerase adds up to 200 adenosines to the 3' end of the RNA transcript. The RNAP2 will continue transcribing, often another several hundred nucleotides, and remain associated with the pre-mRNA that is cleaved as the poly(A) tail is added. As the poly(A) binding proteins disassociate with RNAP2 and the processed mRNA is released, RNAP2 undergoes a change that allows it to disassociate with the DNA and the unused excess RNA transcript is rapidly degraded (*94, 298, 331-333*).

What is certainly appreciated is that the polyadenylation complex begins to associate with the RNAP2 CTD shortly after RNAP2 leaves the transcription start site so the machinery is poised to terminate transcription at any point. Also, the RNA endonuclease that cuts the nascent transcript appears to recognize a wide variety of RNA substrates in a manner only loosely dependent on RNA sequence, allowing for a large variety in the termination sites throughout the eukaryotic transcriptome.

CHAPTER SIX: Methods

6.1 POLYMERASE CHAIN REACTION

To detect RNA, I used reverse transcriptase polymerase chain reaction (PCR) amplification (*273-275*). Treatment of RNA with reverse transcriptase creates a cDNA copy (*87, 276, 277*). PCR uses sequential rounds of heating to 95 ºC to melt DNA, cooling to 60 ºC to anneal sequence specific primers, and then allow primers to be extended by Taq DNA polymerase at 72 ºC (Hotstar Taq, Qiagen). Forward and reverse primers ensure than 2x copies are produced every round of PCR. Thus the abundance of double stranded DNA product increases exponentially to levels that can be detected by electrophoresis in a 1.5% agarose gel and staining with 7 μg / 100 mL ethidium bromide. To detect DNA, I used PCR without the reverse transcriptase treatment.

6.2 PURIFICATION OF POLYADENYLATED RNA

I purified polyadenylated RNA from T47D cells using using a Oligotex® mRNA mini kit (Qiagen). This kit uses oligodT conjugated agarose beads to anneal and precipitate polyadenylated RNA. Beads are washed using spin columns. RNA is eluted from beads by elution buffer heated to 75 ºC. Polyadenylated RNA is a very small fraction of total RNA so concentration could not be measured by UV spectroscopy and had to be estimated from the amount of total RNA used assuming 100% recovery. Based on quantitative real time PCR results, recovery may have been as poor as 25%. Polyadenylated RNA or cDNA, presumably due to its very low concentration, did not store well even at -80 ºC and had to be used within 1 week of its preparation.

6.3 QUANTITATIVE REAL-TIME PCR

In order to measure the relative abundance of RNA, I used quantitative real time PCR (qRT-PCR). For this procedure, PCR is performed with an intercalating dye, Sybr green (Qiagen), to measure the growing abundance of PCR product. The relative abundance of cDNA template is measured using a $\Delta\Delta C_T$ method (*279*). C_T is defined as the cycle at which fluorescence for a PCR reaction rises above an arbitrary threshold. C_T values are normalized to a house keeping gene for a loading control, GapDH (*280*). Also to normalize for relative differences in primer efficiency, I normalized to either genomic DNA or plasmid containing the PR mRNA (K. Horowitz) (*274, 334*). qRT-PCR was performed on an ABI 7500 Real Time PCR System (Applied Biosystems Inc).

6.4 5' AND 3' RAPID AMPLIFICATION OF CDNA ENDS (RACE)

I use the GeneRACER® kit from Invitrogen. For 5' RACE, to eliminate the possibility of detecting degraded RNA products lacking in a 5' cap, I first treated 5 μg of RNA with calf intestinal phosphatase (CIP) at 50 ºC for 1 hour to remove exposed 5' phosphates. After this RNA was recovered by phenol:chloroform extraction and ethanol precipitation incubating for 2 hours at -80 ºC. Then the 5' cap structure was removed from RNA to expose the 5' phosphate by treatment with tobacco acid pyrophosphatase (TAP) at 37 ºC for 1 hour. After this RNA was recovered by phenol:chloroform extraction and ethanol precipitation. Before ligation, RNA was incubated at 65 ºC for 5 minutes to relax RNA secondary structure. A 44 nucleotide RNA linker sequence was ligated to the expose 5' phosphate using T4 RNA ligase at 37 ºC for 1 hour. RNA was recovered by phenol:chloroform extraction and ethanol precipitation (Appendix A).

Following this, RNA is reverse transcribed using Superscript III® (Invitrogen) reverse transcriptase (RT) with random primers (Applied Biosystems Inc.) at 50 ºC for 1 hour followed by heat inactivation at 70 ºC for 10 minutes. Random primers are used because oligodT primers frequently do not produce full length transcripts as reverse transcriptase falls off the transcript before reaching the 5' end. Following the RT reaction,

RNA is cleaved from the cDNA by treatment with RNase H treatment at 37 ºC for 20 minutes. cDNA was diluted to a concentration 20 ng / μl and stored at -80 ºC. cDNA stored well at -80 ºC for more than 2 years.

For 3' RACE, 5 μg of cDNA is prepared by reverse transcription with Superscript III® at 50 ºC for 1 hour using oligodT (24 dTs) primers with a 5' sequence complementary to 3' and 3' nested primers for amplification. In this manner a known sequence is present on the 3' ends of cDNA transcripts. Following reverse transcription, RNA is cleaved from DNA by RNase H treatment at 37 ºC for 20 minutes.

After preparation of 5' and 3' RACE ready cDNA, I used PCR to amplify cDNA ends. I used a high fidelity proof reading Taq, Platinum Taq™ (Invitrogen), without 3' to 5' exonuclease activity. The high fidelity Taq ensured accurate amplification of cDNA sequences, is less tolerant of mismatches in primer sequences, and was able to amplify long and difficult sequences. The lack of 3' to 5' exonuclease activity allowed for Taq to leave a free A on the 3' ends of PCR products. This free A was needed for subsequent cloning using a TOPO TA cloning® kit (Invitrogen).

Two rounds of PCR were used with "nested" primers that reside within the region amplified by the first round of PCR, to increase specificity (**Fig 15a**). In 5' or 3' RACE, every cDNA transcript has the same sequence complementary to the 5' or 3' targeting primers, which dramatically increases nonspecific amplification. Also, higher Mg^{2+} concentration (2 mM) is used, which allows a higher propensity for primers to anneal mismatched sequences. Finally, contrary to the manufacturer's recommendation, I used 20 nucleotide primers with a melting temperature (T_M) around 60 ºC instead of 30 nucleotide long primers with a T_M around 70 ºC. I did this because at the time I was more familiar with PCR using shorter primers with a T_M around 60 ºC. This further increased the propensity for nonspecific product to be amplified even through two rounds of PCR.

PCR product was visualized by agarose gel electrophoresis with a 3% agarose gel. A higher percentage gel was used to separate what the multiple PCR products produced, some of which turned out to be specific and others not. Bands were cut from the agarose gel and product recovered by centrifugation of gel slices at high speed for 3 minutes in a S.N.A.P.™ spin columns (Invitrogen) with a clean 1.5 mL tube. Recovered

flow through containing the PCR product should be in a volume less than 100 μL. If the gel slice was too large and the volume greater than 100 μL, this usually resulted in the product being too low a concentration for cloning.

PCR product was cloned using a TOPO TA cloning® kit (Invitrogen) into a pCR4-TOPO vector. The pCR4-TOPO vector is ideal for sequencing because it contains primer sites for T7, T3, M13 forward, and M13 reverse sequencing on either side of the insert. Also EcoRI restriction sites immediately outside the insert site allow for a quick check of the size of the inserted product by EcoRI digestion. Finally, the plasmid uses positive selection to ensure empty plasmid is never grown by incorporation of a lethal *E. coli* gene, *ccdB*, which is disrupted by insertion of PCR product.

Topoisomerase based cloning works by using linearized plasmid with *Vaccinia* virus topoisomerase I covalently bound to the overhanging 3' deoxythymidine (T) residues through a tyrosyl residue (Tyr-274). The 3' phosphate is then attacked by the 5' hydroxyl of the DNA insert, reversing the reaction and releasing topoisomerase I. I used half the manufacturer's recommended reaction size to conserve reagent. For each reaction I used 0.5 μl TOPO® vector, 0.5 μL salt solution, and 2 μL purified PCR product. Ligated vector was transformed into competent DH15α cells (gift from the M. Rosen lab) and spread on LB-agarose plates containing 100 mg/L ampicillin (Sigma-Aldrich).

Following at least 12 hours of growth at 37 ºC, colonies were selected and transferred to a 10 mL LB (Appendix A) liquid broth and grown overnight. Plasmid was purified using an alkaline lysis miniprep protocol (Current Protocols in Molecular Biology 1.6). Plasmid pellets were resuspended in 30 μL of nuclease free water. Insertion of correct PCR product was verified by EcoRI digestion and gel electrophoresis. Then 1 μL of plasmid was mixed with 10 μL of nuclease free water and 1 μL of RNase A (40 U) and sent to the McDermott Sequencing Core for sequencing using either T7 or M13 reverse primers. The amount of plasmid should be around 1.5 μg as estimated by gel electrophoresis. Concentrations measured by UV spectroscopy tended to be considerably higher, possibly due to RNA contamination.

Sequences were aligned with the human genome using blastn (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Recovered sequences contained repeat elements

that could not be aligned unless the Repeat Masker was turned off under advanced settings.

6.5 NORTHERN ANALYSIS

5 μg of total RNA was run on a denaturing gel (1.0% agarose, 1x denaturing gel buffer, NorthernMax® kit, Ambion) and transferred to GeneScreen Plus ® (PerkinElmer) by vacuum pressure and UV crosslinked. A radiolabeled probe was generated to by asymmetric PCR (Strip-Ez® kit, Ambion) using only a forward primer for 60 cycles with 60 μCi [α^{32}P]dATP (Molecular Probes) and a template generated by PCR amplification of exon boundary 7/8 of the PR mRNA (*335*). This probe was incubated with blots overnight and blots were exposed to film for 3 days to show PR mRNA

6.6 BIOTIN PULLDOWN

MCF7 or T47D were grown in six-well dishes and transfected with biotinylated RNA heteroduplexes at a concentration of 100 nM. Avidin-coated beads were prepared by pre-blocking with yeast tRNA and salmon sperm DNA. Three days following transfection, cells were harvested to obtain nuclei. The nuclei were mixed with avidin-coated beads at 4°C for two hours. S. Younger washed the beads exhaustively prior to elution of RNA using buffer (1.5 % Biotin, 4 M guanidine thiocyanate, 25 mM sodium citrate, 0.5% sodium N-lauroyl sarcosinate) for 2 hours at 45°C with periodic gentle agitation. Samples were treated with DNaseI to remove any contaminating DNA and then amplified by PCR using either primer sets capable of amplifying the target antisense transcript or targeting genomic DNA. The biotin pulldown assay was modified from that reported by Kraynack and Baker (*336*) and the elution buffer was based on that described by Hayashizaki and colleagues (*337*).

6.7 CHROMATIN IMMUNOPRECIPITATION (CHIP)

For this experiment I used chromatin immunoprecipitation of RNA polymerase II (RNAP2). I seeded either 4 million MCF7 cells or 3 million T47D cells into two 15 cm diameter dishes per each treatment. The seeding number varied depending on how fast the cells were growing which varied from lot to lot and by passage number.

Cells were checked for even spreading before transfection. Uneven spreading affected transfection efficiency so negatively that unevenly seeded dishes were discarded and seeding redone. Cells were transfected with 25 nM PR-9, PR-11 or MM on day 2, two dishes per treatment. Media was changed 24 hours later. Cells were harvested on day 5. For harvesting, cells were washed in PBS and a 5 cm^2 sample was taken by scraping and placed in a 1.5 mL tube for either western or Trizol® treatment to recover mRNA. Then 10 mL of PBS with 1% formamide was added and incubated for 10 minutes. Crosslinking was stopped by addition of 1.25 M glycine dissolved in PBS to a final concentration of 125 mM for 5 minutes. Formamide and glycine was aspirated off and 5 mL of PBS was added to the dish. Cells were harvested by scraping with a rubber policeman and the two dishes with the same treatment were combined into one 15 mL conical tube.

Cells were pelleted by centrifugation in a cold centrifuge at 500g (~1700 rpm) for 5 minutes. PBS was aspirated off and cells were resuspended with gentle, pulsed vortexing and adding hypotonic lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, and 0.5% NP40) 1 mL at a time up to 4 mL. Nuclei were pelleted again by centrifugation in a cold centrifuge at 500 g for 5 minutes and hypotonic lysis was repeated. Final pelleted nuclei were resuspended in 1 mL lysis Buffer B (1% SDS, 10 mM EDTA, 50 mM Tris-HCL pH 8.1) with 1x EDTA-free Complete mini® protease inhibitors (Roche) added just before use. Nuclei were incubated with lysis Buffer B for 10 minutes on ice before being stored in -80 ºC.

Cells were sonicated with a Biologics Model 150 V/T ultrasonic homogenizer (T. Kodadek lab). As with any sonicator, conditions for sonication were optimized. For this machine, a series of 30 second sonications were performed on untreated crosslinked

nuclei at 40% power with 1 minute cool down between each sonication. If samples foamed, they could not be sonicated further until allowed to set at -20 ºC overnight for the foam to dissipate. Use of 15 mL conical tubes and keeping the probe all the way at the bottom of the tube reduced the chances of foaming. Sonication was performed with conical tube in an ice bath slurry. After each sonication, 10 µL was taken. To each sample were added 5 µL of 5 M NaCl and 85 µL of water. Samples were incubated at 65 ºC overnight.

The next day, to each sample were added 2 µL of 0.5 M EDTA, 4 µL Tris-HCL (pH 6.5), 1 µL RNase A (40 U), and 1 µL proteinase K (20 µg). Samples were incubated at 40 ºC for 1 hour. Samples were then extracted with 100 µL phenol:chloroform (Fluka) and spun at maximum speed for 13 minutes. The aqueous layer was taken and to each sample was added 10 µL Na-acetate (pH 5.2) and 220 µL ethanol. Samples were allowed to precipitate at -80 ºC for at least 2 hours. DNA was pelleted by centrifugation at maximum speed for 13 minutes. Ethanol was aspirated and the samples resuspended in 100 µL of water and 10 µL were run on a 1.5% agarose gel stained with ethidium bromide. Viewing the gel revealed that after two 30 second pulses at 40% power, most of the DNA was sheared between 500 and 1000 nucleotides. The ability for ChIP to resolve one binding site from another does not get much if any better than 500 nucleotides so further sonication only results in loss of signal.

Treated samples were sonicated for two 30 second pulses at 40% power. After sonication samples were transferred to 1.5 mL tubes and centrifuged at maximum speed for 15 minutes. The supernatant ("lysate") was transferred to new 1.5 mL tubes and stored at -80 ºC. ChIP lysates keep for up to one year but around one year later the efficiency of pulldown seemed reduced and much of the chromatin appeared to have precipitated out of solution. Perhaps brief sonication would correct this problem but I have not verified this.

The high level of SDS is important to keep chromatin in suspension. However, these high levels would prevent antibodies from interacting with their protein antigens. For this reason, 100 µL of lysate in diluted in 900 µL of IP Buffer (0.01% SDS, 1.1% Triton-X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl) with 1x

Complete mini® protease inhibitor cocktail (Roche) added just before use. To this is added either 2 μg of anti-RNAP2 antibody clone CTD4H8 (Millipore) or 2 μg of mouse IgG for negative control. Lysate and IP buffer with incubated with antibody at 4 ºC overnight with rotation.

The next day 60 or 70 μL of protein G Plus/ protein A™ agarose beads (washed and resuspended in IP Buffer, IP05 Calbiochem) are added to each sample and incubated at 4 ºC with rotation for at least 2 hours. Beads are recovered by centrifugation at 1500 rpm for 2 minutes. Supernatant is carefully pipetted off using a 1 μL pipette followed by a 200 μL pipette. A gel loading 200 μL tip or standard 200 μL tips can be used. I pipette the waste into either down the side of a 50 mL conical or a glass beaker so that I can monitor the amount of beads lost as they flow down the side and I can see the total bead loss after all the washes. Care must be taken to keep total bead loss to less than 20% through all the washes.

Beads are kept on ice and resuspended in cold washes. Washes are mixed by hand every 2 minutes for a total of 6 minutes and then beads are pelleted by spinning at 2000 rpm for 5 minutes, washes removed, and new wash added. Five washes are performed with the fifth wash being optional. The first wash is low salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl). The second is high salt wash, which is low salt wash but with 500 mM NaCl. The third wash is a LiCl wash (0.25M LiCl, 1% NP40, 1% deoxycholate w/v, 1 mM EDTA, 10 mM Tris-HCl pH 8.1). The last two washes are simply Tris-EDTA pH 8.0 (Ambion).

Following washing, protein is eluted off beads at room temperature for 30 minutes with 400 μL of Elution Buffer (1% SDS, 0.1M $NaHCO_3$, in water). $NaHCO_3$ is stored as 1M stocks in -20 ºC and Elution Buffer is made fresh. 20 μL of lysate is added to 380 μL elution buffer for "input" samples. Following elution 16 μL of 5M NaCl is added to a final concentration of 200 mM and samples are incubated at 65 ºC for at least 2 hours to reverse the formaldehyde crosslinking. DNA can be incubated overnight but not RNA. At any step during washing, the wash can be removed and the beads stored at -80 ºC overnight. Elute before or after reversing crosslinking can be stored at -80 ºC overnight.

After reversing crosslinking, to each sample is added 32 μL of 1M Tris-HCl pH 6.5, 16 μL of 0.5M EDTA pH 8.0, 1 μL RNaseA, and 1 μL proteinase K. Samples are incubated at 40 ºC for 1 hour. Following proteinase K (Invitrogen) digestion, 400 μL of phenol:chloroform are added, samples are shaken, and centrifuged at maximum speed for 15 minutes. The aqueous layer is transferred to a new 1.5 mL tube and to each sample is added 2 μL glycogen (Sigma-aldrich), 40 μL Na-acetate pH 5.2, and 880 μL of ethanol. Samples are shaken and allowed to precipitate for at least 2 hours at -80 ºC.

Following precipitation samples are centrifuged for 15 minutes at maximum speed. Ethanol is carefully removed by pipetting. Pellets washed in at least 700 μL of 70% ethanol and centrifuged for 5 minutes at maximum speed. Wash is carefully removed by pipetting and pellets are resuspended in 50 μL of nuclease free water. Samples can be stored indefinitely at either -20 ºC or at -80 ºC. For quantification of relative abundance of RNAP2 by qRT-PCR, 1 or 2 μL of DNA is added to each PCR reaction depending on the efficiency of precipitation.

For RNAP2, primers targeting the gene GapDH can also be used to normalize for sample loss but this does not work for other antibodies. All samples are normalized to their respective inputs. IgG negative controls are measured as a fraction of their relative specific antibody pulldown. For ChIP experiments, inputs should have C_T's in the mid to low 20's, IgG's should have C_T's higher than 31 or 32, and samples of interest should have C_T's between the mid 32 and the mid 20's. If samples come up at or below IgG negative control samples then the antibody is only pulling down random associations. Some samples, such as RNAP2 for a very highly expressed gene like beta-actin, will have C_T's at or very near the inputs.

6.8 RNA IMMUNOPRECIPITATION (RIP)

This RIP protocol is nearly identical to the ChIP protocol with a few considerations for purifying RNA rather than DNA. The primary difference is that RNase inhibitor must be used to protect RNA from degradation. The yield of RIP product is very low so more material is needed. I seeded either 4 million MCF7 cells into four 15 cm

diameter dishes per treatment or 3 million T47D cells into six 15 cm diameter dishes per each treatment. The seeding number varied depending on how fast the cells were growing which varied from lot to lot and by passage number.

As with ChIP, cells were checked for even spreading before transfection. Unevenly seeded dishes were discarded and seeding redone. Cells were transfected with 25 nM PR-9, PR-11 or MM on day 2, two dishes per treatment. Media was changed 24 hours later. Cells were harvested on day 5. For harvesting, cells were washed in PBS and a 1.5 in^2 sample was taken by scraping and placed in a 1.5 mL tube for either western or Trizol® treatment to recover mRNA. Then 10 mL of PBS with 1% formamide was added and incubated for 10 minutes. Crosslinking was stopped by addition of 1.25 M glycine dissolved in PBS to a final concentration of 125 mM for 5 minutes. Formamide and glycine was aspirated off and 5 mL of PBS was added to the dish. Cells were harvested by scraping with a rubber policeman and the dishes with the same treatment were combined into one 15 mL conical tube.

Cells were pelleted by centrifugation in a cold centrifuge at 500g (~1700 rpm) for 5 minutes. PBS was aspirated off and cells were resuspended with gentle, pulsed vortexing and adding hypotonic lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, and 0.5% NP40) 1 mL at a time up to 4 mL. Nuclei were pelleted again by centrifugation in a cold centrifuge at 500 g for 5 minutes and hypotonic lysis was repeated. Final pelleted nuclei were resuspended in 1 mL lysis Buffer B (1% SDS, 10 mM EDTA, 50 mM Tris-HCL pH 8.1) with 1x EDTA-free Complete mini® protease inhibitors (Roche) and RNaseIN (Promega) RNA inhibitors added just before use. Nuclei were incubated with lysis Buffer B for 10 minutes on ice before being stored in -80 ºC.

For my RIPs, samples were not sonicated. However, recent experiments by Y. Chu have suggested that sonication may slightly increase signal recovery. Samples were transferred to 1.5 mL tubes and centrifuged at maximum speed for 15 minutes. The supernatant ("lysate") was transferred to new 1.5 mL tubes and stored at -80 ºC. RIP lysates keep for up to one year.

100 μL of lysate is diluted in 900 μL of IP Buffer (0.01% SDS, 1.1% Triton-X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl) with 1x Compete mini®

protease inhibitor cocktail (Roche) and RNaseIN RNA inhibitors added just before use. To this is added either 2 μg or 4 μg of specific antibody or mouse IgG for negative control. Lysate and IP buffer is incubated with antibody at 4 ºC overnight with rotation.

The next day 60 or 70 μL of protein G Plus/ protein A equalibriated agarose beads (washed and resuspended in IP Buffer, Calbiochem) are added to each sample and incubated at 4 ºC with rotation for at least 2 hours. Beads are recovered by centrifugation at 1500 rpm for 2 minutes. Supernatant is carefully pipetted off using a 1 μL pipette followed by a 200 μL pipette. A gel loading 200 μL tip or standard 200 μL tips can be used. I pipette the waste into either down the side of a 50 mL conical or a glass beaker so that I can monitor the amount of beads lost as they flow down the side and I can see the total bead loss after all the washes. Care must be taken to keep total bead loss to less than 20% through all the washes.

Beads are kept on ice and resuspended in cold washes. Washes are mixed by hand every 2 minutes for a total of 6 minutes and then beads are pelleted by spinning at 2000 rpm for 5 minutes, washes removed, and new wash added. Five washes are performed with the fifth wash being optional. The first wash is low salt (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl). The second is high salt wash, which is low salt wash but with 500 mM NaCl. The third wash is a LiCl wash (0.25M LiCl, 1% NP40, 1% deoxycholate w/v, 1 mM EDTA, 10 mM Tris-HCl pH 8.1). The last two washes are simply Tris-EDTA pH 8.0 (Ambion). I do not add RNase inhibitors to washes and do not see significant degradation. All washes have EDTA and high SDS concentrations, which should help inhibit RNases.

Following washing, protein is eluted off beads at room temperature for 30 minutes with 400 μL of Elution Buffer (1% SDS, 0.1M $NaHCO_3$, in water). $NaHCO_3$ is stored as 1M stocks in -20 ºC and Elution Buffer is made fresh. 20 μL of lysate is added to 380 μL elution buffer for "input" samples. Following elution 16 μL of 5M NaCl is added to a final concentration of 200 mM and samples are incubated at 65 ºC for at least 2 hours to reverse the formaldehyde crosslinking. DNA can be incubated overnight but not RNA. At any step during washing, the wash can be removed and the beads stored at -

80 ºC overnight. Elute before or after reversing crosslinking can be stored at -80 ºC overnight.

After reversing crosslinking, to each sample is added 32 µL of 1M Tris-HCl pH 6.5, 16 µL of 0.5M EDTA pH 8.0, 1 µL RNaseIN RNase inhibitor, and 1 µL proteinase K. Samples are incubated at 40 ºC for 1 hour. Following proteinase K (Invitrogen) digestion, 400 µL of phenol:chloroform are added, samples are shaken, and centrifuged at maximum speed for 15 minutes. The aqueous layer is transferred to a new 1.5 mL tube and to each sample is added 2 µL glycogen (Sigma-aldrich), 40 µL Na-acetate pH 5.2, and 880 µL of ethanol. Samples are shaken and allowed to precipitate for at least 2 hours at -80 ºC.

Following precipitation samples are centrifuged for 15 minutes at maximum speed. Ethanol is carefully removed by pipetting. Pellets washed in at least 700 µL of 70% ethanol and centrifuged for 5 minutes at maximum speed. Wash is carefully removed by pipetting and pellets are resuspended in 20 µL of nuclease free water. RIP samples do not store well, likely due to the low concentration. At best, RNA or cDNA can be effectively stored at -80 ºC for up to two weeks. 10 µL of RNA is treated with DNase (Worthington) and used for cDNA synthesis.

For quantification of relative abundance of RNA by qRT-PCR, 2 or 4 µL of DNA is added to each PCR reaction depending on the efficiency of immunoprecipitation. Even though qRT-PCR is used to quantify RIP product, the resulting C_T's are usually in the high 30's and do not appear to amplify linearly. For this reason, I recommend displaying RIP data using gel electrophoresis rather than qRT-PCR results. To make publication quality gels, 7 to 10 µL of the qRT-PCR reaction is run on a 3 or 4% agarose gel stained with ethidium bromide.

6.9 CHROMATIN CONFORMATION CAPTURE (3C)

The cell culture and harvest of 3C material is very similar to that for ChIP. After the harvest alloquotes of 1 million nuclei in 100 µL of 0.125 M glycine-PBS are stored at

-80 ºC. Since it is difficult to estimate 1 million cells, I use one alloquot of uncrosslinked nuclei and purify genomic DNA. 1 million cells should have approximately 5 μg of genomic DNA. I measure the amount of genomic DNA in one uncrosslinked aloquot and then divide the aloquots before each 3C experiment so that 5 μg of genomic DNA is used in each experiment. My 3C protocol is derived from protocols published by the Baylin, Proudfoot, Dekker, and Ohlsson labs (*294, 305, 338, 339*). The basic outline for a 3C experiment is 1) formaldehyde crosslink, 2) restriction enzyme cleave DNA, 3) ligate ends held in proximity by protein interactions by T4 DNA ligation in a dilute solution, 4) purify ligated DNA product, and 5) detect ligated product by PCR amplification using primers sets that uniquely amplify ligated product rather than undigested DNA.

I seeded either 4 million MCF7 cells or 3 million T47D cells into two 15 cm diameter dishes per each treatment. The seeding number varied depending on how fast the cells were growing which varied from lot to lot and by passage number. I seed two dishes per each treatment because one will be formaldehyde crosslinked and one will be a no formaldehyde crosslink control.

Cells were checked for even spreading before transfection. Uneven spreading effected transfection efficiency so negatively that unevenly seeded dishes were discarded and seeding redone. Cells were transfected with 25 nM PR-9, PR-11 or MM on day 2, two dishes per treatment. Media was changed 24 hours later. Cells were harvested on day 5. For harvesting, cells were washed in PBS and a 5 cm^2 sample was taken by scraping and placed in a 1.5 mL tube for either western or Trizol® treatment to recover mRNA. Then, for crosslinked samples, 10 mL of PBS with 1% formamide was added and incubated for 10 minutes. For no crosslinking control, 10 mL of PBS without formaldehyde is added. Crosslinking was stopped by addition of 1.25 M glycine dissolved in PBS to a final concentration of 125 mM for 5 minutes. Formamide and glycine was aspirated off and 5 mL of PBS was added to the dish. Cells were harvested by scraping with a rubber policeman and the two dishes with the same treatment were combined into one 15 mL conical tube.

Cells were pelleted by centrifugation in a cold centrifuge at 500g (~1700 rpm) for 5 minutes. PBS was aspirated off and cells were resuspended with gentle, pulsed

vortexing and adding hypotonic lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, and 0.5% NP40) 1 mL at a time up to 4 mL. Nuclei were pelleted again by centrifugation in a cold centrifuge at 500 g for 5 minutes and hypotonic lysis was repeated. Nuclei are finally resuspended in 1 mL of 0.125 M glycine-PBS and separated into 100 μL alloquotes. One no crosslink alloquot is kept aside for purifying and quantifying genomic DNA.

After quantifying genomic DNA in the alloquots, an appropriate number of alloquots is removed and thawed for experiments. Crosslinked and uncrosslinked nuclei are diluted to 5 μg of genomic DNA per sample in 100 μL of 0.125 M glycine-PBS kept in 1.7 mL microcentrifuge tubes. I designed my experiments so that the ligation step could be performed in 2.2 mL volumes using 2.2 mL microcentrifuge tubes. Samples are diluted to 500 μL in 1.7 mL microcentrifuge tubes in 1.2x restriction enzyme buffer with 0.3% SDS and incubated with shaking at 37 ºC for 1.5 hours. This is supposed to "loosen" crosslinked chromatin to allow access for restriction enzymes. However, digestion is not near 100% efficient and some digest sites are not effectively cleaved, likely due to the inhibition from formaldehyde crosslinking. SDS is sequestered after 1.5 hours by adding 108 μL of 10% Triton-X (1.8% final concentration) with shaking at 37 ºC for 1 hour.

I designed the 3C experiment to use the restriction enzyme DpnII. I chose this restriction enzyme because it was used by Tan-Wong et al so I could exactly reproduce their experiment for BRCA1. Also, I mapped out every DpnII cut site for the PR gene locus and found several cut sites at the 5' and 3' ends of the PR gene suggesting this would be an appropriate restriction enzyme to study this locus as well. I screened several potential restriction enzymes using Restriction Mapper ver. 3 to generate maps and label every cut site in my region of interest (http://www.restrictionmapper.org). Other restriction enzymes screened did not offer cut sites at both the 5' and 3' of the PR locus.

Several requirements must be met for a useful restriction enzyme. 1) It is desireable that they have one or more cut sites at both regions of interest and several cut sites within and outside those regions. For this reason, useful restriction enzymes must balance between being overly promiscuous cutters and being promiscuous enough. 2)

Restriction enzymes should be either sticky end or blunt end cutters. Class II or asymmetric restriction enzymes cannot be used. I chose to use sticky end cutters. However, Lieberman-Aiden et al recently published a genome wide 3C study using blunt end ligation (*340*). 3) I used restriction enzymes that cut optimally at 37 ºC. Heating samples above 45 ºC should be avoided to prevent reversal of formaldehyle crosslinking. On the other hand, restriction enzymes that cut effectively at low temperatures may allow some restriction enzymes that are not inactivated to continue cutting during the T4 DNA ligation step at 16 ºC. 4) As is commonly used by the Proudfoot and Baylin labs, it would be desireable that the 3C experiment be designed so that more than one restriction enzyme is used to hopefully give convergent results that reinforce each other. I did not do this due to issues of time and we had an alternative assay, RIP, that gave a similar supporting result.

After sequestering SDS with Triton-X, 200 to 300 units of restriction enzyme are added and samples are incubated at 37 ºC overnight with shaking. If desired, restriction enzyme digest can be performed with a total volume of 700 μL and 100 μL can be removed from samples after digestion and crosslinking reversed to test with PCR primers designed to test the cleavage efficiency of the restriction enzyme. For testing a new enzyme, this may be useful in case no looping is detected to ensure that cleavage did occur. Lack of cleavage may be due to over extensive formaldehyde crosslinking or that the restriction enzyme may be sensitive to the high salt and detergent concentrations in the buffer. As mentioned, cleavage efficiency is definitely less than 100% so qRT-PCR should be used to measure efficiency.

On Day 2, samples are removed from shaking. 112 mL of 10% SDS is added to a final concentration of 1.6% SDS to inactivate the restriction enzyme. Samples are incubated with SDS for 1 hour at 37 ºC. After this, samples are switched to 2.2 mL microcentrifuge tubes and samples brought up to 2 mL volume with 1.2x final concentration of T4 DNA ligase buffer (approximately 2.5 ng/μL final concentration of DNA (*305*)). To this is added 200 mL of 10% Triton-X to sequester the SDS and samples are incubated at 37 ºC for 1 hour.

Final concentration of DNA in the dilute T4 reaction is important to prevent non-specific ligation of uncrosslinked DNA ends. Ultimately the no crosslink control shows if non-specific amplification is occurring. Some protocols dilute more heavily, up to 6 or more mL (*339*). For this reason, some protocols allow ligation to progress for up to 3 to 9 days (*338*). Large volumes are difficult to manage, especially when phenol:chloroform extracting the DNA. Also, it is doubtful that T4 DNA ligase would remain enzymatically active for much more than 8 hours, making several day incubations seem unfruitful (for more information see Invitrogen or NEB product insert for T4 DNA Ligase).

Samples are then cooled on ice and 40 units of T4 DNA ligase is added (NEB). Samples are incubated in a 16 ºC water bath overnight. To get a water bath to 16 ºC, use a 4 ºC fridge or the cold room and add ice to help cool water bath faster. Use a thermometer to make sure the temperature is actually 16 ºC in the bath as the temperature will indicate something lower such as 14.6 ºC. For a 4 hour T4 DNA ligation reaction, it may be sufficient to prepare an ice/water slurry at the bench top with a thermometer to achieve 4 hours of 16 ºC. After overnight incubation I allow T4 DNA Ligase to continue for 30 minutes at room temperature (*305*) even though the product insert states that 20 minutes at room temperature is likely enough to kill the T4 DNA Ligase enzyme.

To reverse formaldehyde crosslinking, 80 µL of 5M NaCl and 20 µg of Proteinase K is added and samples are incubated at 65 ºC for 2 hours. Following reversal of crosslinking, samples are split into two 2.2 mL microcentrifuge tubes for phenol:chloroform extractions followed by isopropanol precipitation. Equal volume of phenol:chloroform:isomyl alchohol are added to each half of the sample (1.1 mL) and shaken then centrifuged at more than 11000 rpm for 15 minutes. Then the top aqueous layer is placed in a new tube and to that is added $1/10^{th}$ volume Na-acetate (110 µL), 1.5 µL glycogen, and 1 mL of isopropanol. Samples are incubated for at least 2 hours at -80 ºC.

Following precipitation, samples are thawed and centrifuged at more than 11000 rpm for 15 minutes. Supernatent is carefully poured off so not to dislodge the pellet and 750 µL of 75% ethanol in nuclease free water is added. Pellets are centrifuged at 11000

rpm for 7 minutes. Wash is pipetted off carefully and final ethanol is removed using a p200 pipette. Pellets are resuspended and combined in 50 μL of nuclease free water.

Primers for PCR detection were designed using primer3 software (http://frodo.wi.mit.edu/primer3/). Primers were designed to bracket each restriction enzyme site of interest leaving around 40 to 70 nucleotides of distance between the primer 5' end and the cut site (i.e. including the primer sequence). This is to ensure that ligated 3C PCR product can be amplified with products between 60 and 150 nucleotides long. Multiple primers sets were designed to certain key cut sites of interest. For detection, either the forward or reverse primer from one cut site of interest is combined with the forward or reverse primer from the other cut sites of interest. Since restriction enzymes cut palendromic sequences, they can be ligated in either orientation so it does not matter whether two forward, two reverse, or one forward and one reverse primer is used for detection. Some cut sites fell in repeat elements which turned out to be extremely problematic leading to robust amplification of nonspecific product.

To detect 3C product, I have used a combination of qRT-PCR and standard PCR followed by gel electrophoresis. For my first 3C experiments, some contaminant, likely SDS, seemed to carry over and inhibit the PCR reaction resulting in high C_T's well outside of the linear range for qRT-PCR. My last 3C experiments, the products were much more efficiently amplified to yield C_T's in the high to low 20's. For these experiments, I felt comfortable quantifying results by qRT-PCR. I performed serial dilutions of 3C product using both qRT-PCR and semi-quantitative PCR where samples are quantified by density of the bands with ethidium bromide staining to test that amplification was linear. Product was not amplified very linearly with an efficiency of between 120% and 150% likely due to the abundance of nonspecific products amplified. However, this efficiency is sufficient to make order of magnitude approximations of relative concentration.

Products were visualized by gel electrophoresis on a 3% agarose gel stained with ethidium bromide. After at least 35 and sometimes up to 45 rounds of PCR, a solid band of the predicted molecular weight could be visualized along with some faint nonspecific bands. These bands are easily seen with fewer rounds of PCR and with less background,

but they are often not bold enough for publication. Sometimes a band of an alternate molecular weight could be visualized due to use of a nearby alternative restriction cut site. All products were purified from the agarose gel, cloned, and sequenced to ensure detection of correct product. Some products were of the correct molecular weight but proved to be nonspecific. Some products were not of the expected molecular weight but were specific resulting from inefficient restriction enzyme digestion and inclusion of more than one restriction enzyme cut site in the amplified PCR product.

For no-crosslink controls, more than 45 rounds of PCR will amplify something nonspecifically and sometimes of a molecular weight near that of the specific product. For this reason, to make a gel to show no-crosslink controls, it may be desireable to use fewer rounds of PCR, such as 30 or 35 rounds. All amplified products should be sequenced to discover what is being amplified and not assumed based on the molecular weight.

CHAPTER SEVEN: Other studies

7.1 TARGETING INTRONS WITH agRNAS

An attractive model for agRNA activity is that agRNAs may bind the transcript AT-2 while still nascent before splicing and processing occurs. If this were true, agRNAs should be able to target introns of the transcript to recruit argonaute to the antisense transcript and regulate PR expression. Alternatively, if agRNAs recognize the antisense transcript post-processing which may remains somehow associated with chromatin, it should be possible to target distant exons or splice junctions and thereby regulate PR expression.

To test these possibilities, I designed duplex RNAs that target within the first intron of the antisense transcript, some 89, 138, and 400 nucleotides upstream of the PR transcription start site (In-89, In-138, and In-400). I also designed duplexes targeting distant exons 2 and 6 situated some 800 and 65000 nucleotides upstream of the PR transcription start site (Ex2, Ex6A, Ex6B) and spanning the exon junction of exons 1 and 2 (Ex1-2) and exons 5 and 6 (Ex5-6).

I performed two experiments transfecting these duplexes into T47D cells at 50 nM concentration in 6-well dishes. I measured knockdown of the antisense transcript and PR at the level of mRNA using qRT-PCR. For the first experiment I used mismatched duplexes as a negative control and the silencing agRNA PR-9 and the siRNA PR3593 (a.k.a. S2) targeting the PR mRNA (**Fig 54 top**). Unfortunately, PR3593 for reasons not yet fully explored silences PR protein expression without inducing pronounced silencing of PR mRNA. X. Yue has since produced preliminary evidence that this siRNA may silence without cleaving the mRNA, more akin to the action of miRNAs. For this reason on the second experiment I used mismatched duplexes for negative controls, and the 5' agRNA PR-9 and the 3' agRNA PR13580 silencing duplexes for positive controls (**Fig 54 bottom**).

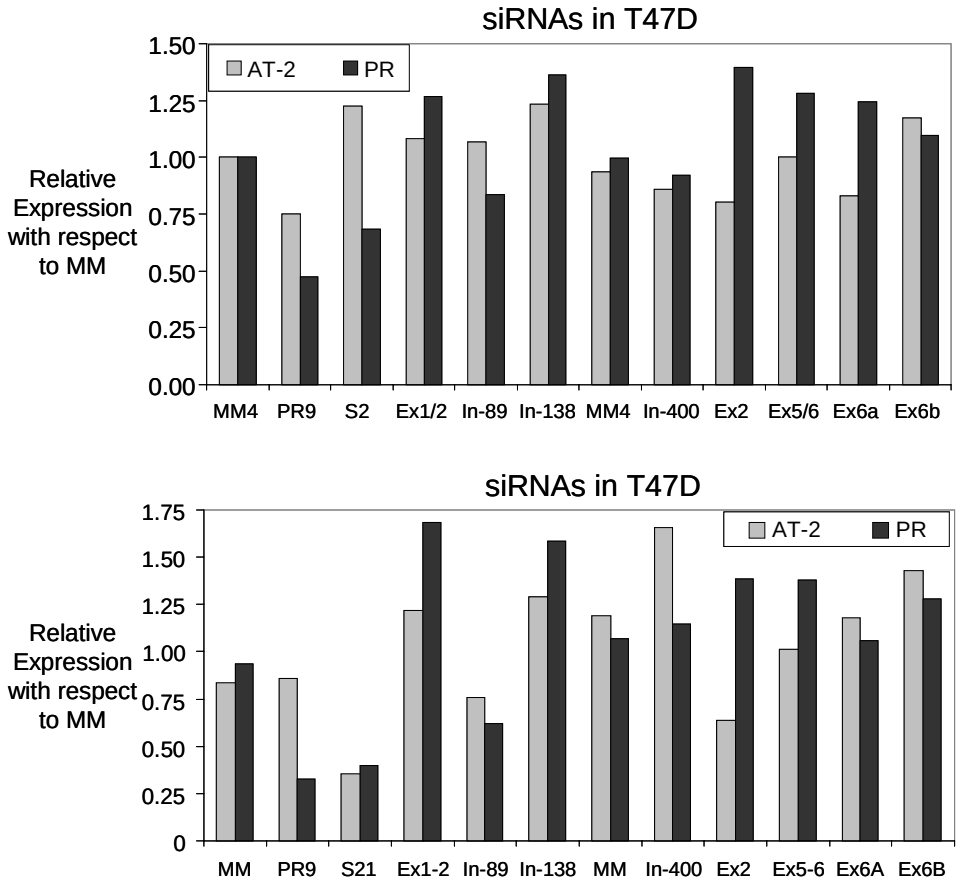


Figure 54: Testing agRNAs targeting AT-2. agRNAs were tested in T47D cells targeting introns, distant exons, and exon boundaries of the antisense transcript AT-2. qRT-PCR was used to measure levels of PR mRNA or AT-2 transcript. Positive control agRNAs include promoter targeting PR-9, (top) mRNA targeting S2 (a.k.a. PR 3593), and (bottom) 3' targeting agRNA S21 (a.k.a. PR13580)

In both experiments, PR-9 did not significantly alter the antisense transcript but PR expression was reduced at the level of mRNA. As mentioned before, PR3593 led to only mediocre reduction in PR expression of around 30% relative to mismatched RNA. PR13580 silenced both PR mRNA and the antisense transcript AT-2 (**Fig 54**). These control experiments contrast slightly from those found by X. Yue in Chapter 5. The reason is likely due to the choice of primers used to measure the antisense transcript AT-2. The primers used in this experiment span exon boundary 2-3 of the antisense transcript and thus measure only mature spliced transcript AT-2. The primers used by X. Yue lie in the PR promoter, able to measure both pre-processed and post-process transcripts and both spliced AT-2, as well as unspliced AT-1 transcripts, whose steady state abundance is around 100 fold lower than that of AT-2.

For each of these experiments, no intron targeting duplex or duplex targeting distant exons had any effect on PR expression. Melting temperatures were taken on the duplexes and showed that duplexes appeared to be annealed properly into a stable duplex. Additionally, no significant changes in the antisense transcript AT-2 levels were observed (**Fig 54**). This negative result may be for several reasons. The duplexes may for some reason be unable to recognize the antisense transcript. Alternatively, the duplexes may bind the antisense transcript but the sites are too distant from the PR promoter to have an effect on PR transcription. The fact the antisense transcript is not silenced does not preclude the possibility of RNA binding. If the antisense transcript does not escape the nucleus, there is not conclusive evidence yet whether argonaute in the nucleus cleaves RNA targets or is somehow post-translationally modified to not allow cleavage.

7.2 OTHER PROTEINS INVOLVED AN agRNA-PROTEIN COMPLEX

In Chapter 4, a possible role for proteins HP1γ and hnRNP-K was investigated in agRNA mechanism (**Fig 26**). It would be helpful for discovery of endogenous agRNAs to know members of the human agRNA protein complex and how that differs from the constitution of the human RISC complex involved in post-transcriptional gene silencing. Unfortunately, HP1γ (a.k.a. chromobox 3 or Cbx3) and hnRNP-K do not follow

recruitment of argonaute proteins for agRNAs targeting the PR promoter eliminating them as promising candidates for members of an agRNA-protein complex.

The *S. pombe* RITS complex may serve as a guide for putative members of an agRNA-protein complex. The member of the RITS complex in *S. pombe*, Chp1, has only one N-terminal conserved protein domain, known as a chromobox domain. Some chromobox domain containing proteins, which may have a role analogous to Chp1 from the RITS complex in *S. pombe*, include Cbx1 (a.k.a. HP1β), Cbx4, Cbx5 (a.k.a. HP1α), Cbx6, Cbx7, and Cbx8. The HP1 proteins (Cbx1, 3, and 5) differ from other chromobox proteins in that they have a C-terminal chromoshadow domain. Changes in HP1 protein recruitment are associated with changes in H3K9 methylation (*341*). Cbx4, 6, 7, and 8 are suggested to be members of the polycomb silencing complexes (*295-297*). Changes in recruitment of polycomb silencing complexes are often associated with changes in H3K27 methylation (*296*). This protein complex has been repeatedly implicated in the activity of noncoding RNAs regulating transcription (see Chap 2).

I have tested for increase in H3K9 trimethylation (Millipore, 17-625) for agRNAs targeting the PR promoter for the case of gene silencing and gene activation. Only marginal upregulation is seen for PR-9 induced gene silencing in T47D cells (n = 3) (**Fig 55a**). No change is seen for PR-11 induced gene activation in MCF7 cells (n = 3) (**Fig 55b**). In the case of MCF7 cells, H3K9 trimethylation specific antibody pulled down the PR promoter considerably higher than negative IgG control suggesting that this cell line may have the PR promoter heterochromatically marked for silencing, but more corroborating lines of evidence would be desirable such as comparing levels of H3K9 methylation in other promoters such as GapDH or β-actin. As mentioned in Chapter 5, increase in H3K27 trimethylation is seen for T47D cells in gene silencing and decrease in H3K27 trimethylation is seen for MCF7 cells in gene activation.

Using ChIP, I have tested for recruitment of HP1α (clone 15.19S2, Millipore 05-689) and β to the PR promoter in the case of agRNA induced gene silencing (n = 4) and activation (n = 1) (**Fig 55e-f**). HP1α did not pull down significantly above negative IgG control (Upstate 12-371), suggesting that HP1α may not associate with the PR promoter. However, this experiment is lacking a positive control to suggest that the HP1α antibody

successfully pulls down using this ChIP protocol. HP1β only showed marginally increased association with the PR promoter in agRNA treated samples (**Fig 55c, f**). Compared to the higher levels of recruitment seen for argonaute and H3K27 trimethylation, a question remains whether these low levels are biologically relevant. An appropriate experiment may be to perform a siRNA knockdown of these proteins, however, their importance in general chromatin stability may lead to a complex response.

Another chromatin marker associated with changes in transcriptional states is histone acetylation. Typically, histone acetylation is associated with increased transcription (*197, 342*). However, as noted in section 1.3, this chromatin marker can cause different transcriptional outputs. For the PR gene, using both 5' and 3' targeting silencing agRNAs in T47D cells, elevates H3 acetylation (**Fig 56b**). Using both 5' and 3' targeting activating agRNA in MCF7 cells, reduce H3 acetylation (**Fig 56a**). These experiments suggest a role for histone acetylases or deacetylases in the switch of transcriptional states initiated by agRNAs. This role may or may not be a ubiquitous feature of agRNA activity, which would be interesting to investigate further.

In section 1.6, we saw evidence suggesting small RNA mediate gene activation involving the protein FXR1. For this reason, I tested if FXR1 may be involved in the agRNA mechanism. I ordered a commercially available anti-FXR1 antibody from Abnova (clone 2G11-00A) and performed a ChIP experiment on MM or PR-9 treated T47D cells (n = 1). Using this antibody, I did not pull down the PR promoter above negative IgG control levels suggesting that the antibody did not pull down DNA. I questioned whether the quality of the antibody may be an issue. I then contacted the G. Dreyfuss lab and acquired the 6BG10 anti-FXR1 antibody used by Vasudevan et al (*82*). I performed ChIP in MCF7 (n = 1) and T47D cells (n = 1) treated with either MM, silencing, or activating agRNA but did not pull down PR promoter above negative IgG control (**Fig 55e-f**). If further experiments were pursued about the involvement of FXR1, it would be interesting to test by western whether the FXR1 protein is in the nucleus or sequestered to the cytoplasm. The effects of siRNA knockdown of FXR1 on agRNA activity may also be interesting. However, these experiments yield no promise for future interesting results.

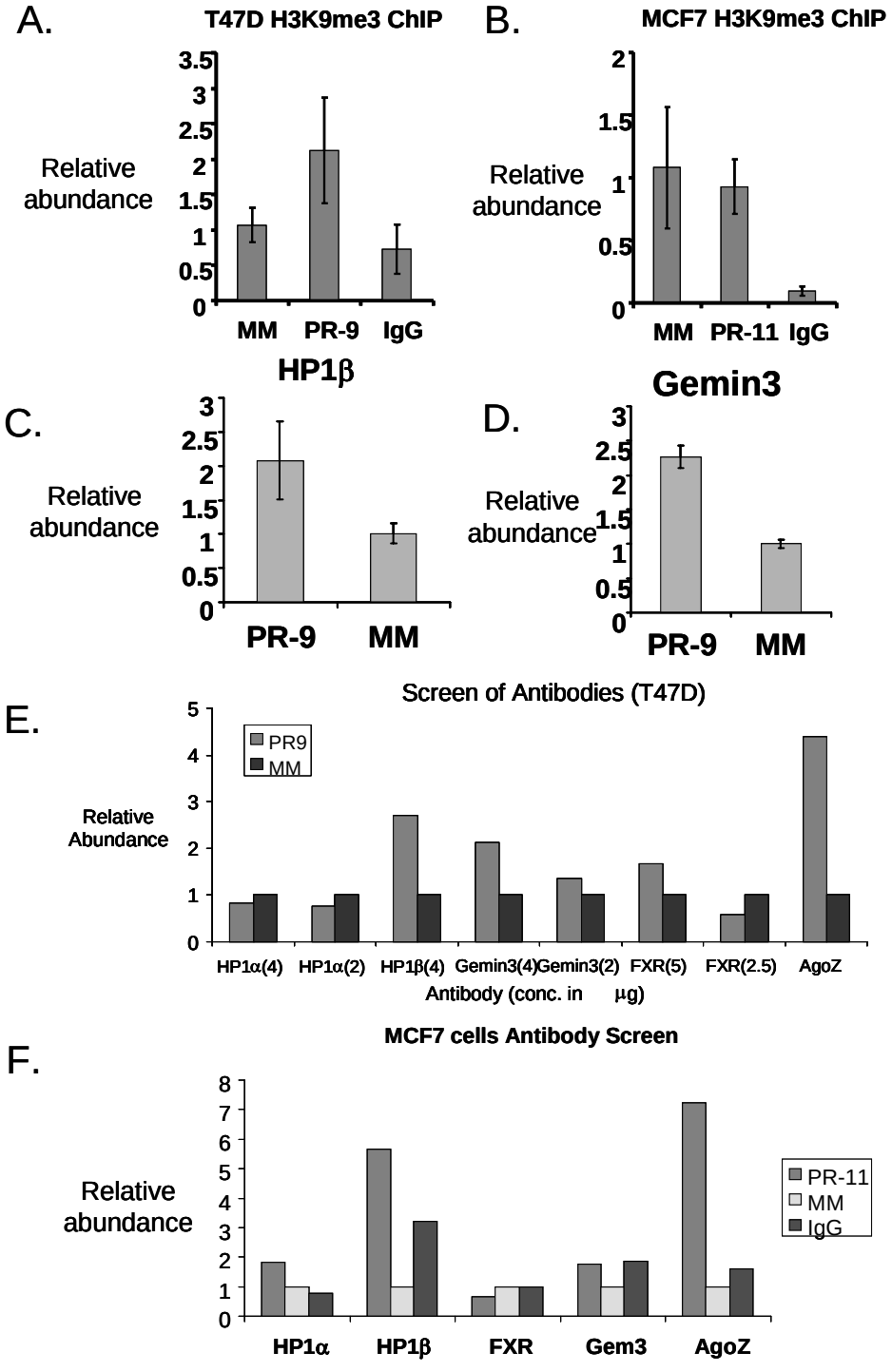

A.
T47D H3K9me3 ChIP
Relative abundance
3.5
3
2.5
2
1.5
1
0.5
0
MM
PR-9
IgG
HP1β
B.
MCF7 H3K9me3 ChIP
Relative abundance
2
1.5
1
0.5
0
MM
PR-11
IgG
C.
Relative abundance
3
2.5
2
1.5
1
0.5
0
PR-9
MM
D.
Gemin3
Relative abundance
3
2.5
2
1.5
1
0.5
0
PR-9
MM
E.
Screen of Antibodies (T47D)
PR9
MM
Relative Abundance
5
4
3
2
1
0
HP1α(4)
HP1α(2)
HP1β(4)
Gemin3(4)
Gemin3(2)
FXR(5)
FXR(2.5)
AgoZ
Antibody (conc. in μg)
F.
MCF7 cells Antibody Screen
Relative abundance
8
7
6
5
4
3
2
1
0
PR-11
MM
IgG
HP1α
HP1β
FXR
Gem3
AgoZ

Figure 55: ChIP screen of protein recruited by agRNAs. H3K9 trimethylation is slightly increased in T47D cells after treatment with silencing agRNA PR-9 (A) but not in MCF7 cells after treatment with activating agRNA PR-11 (B). Silencing agRNA PR-9 in T47D cells slightly elevated recruitment of HP1β (C) and Gemin3 (D) (n = 3). In T47D cells, antibodies targeting HP1α and FXR1 did not pull down above IgG negative control (E). In MCF7 cells, antibodies targeting Gemin3 and FXR1 did not pull down above IgG negative control (F).

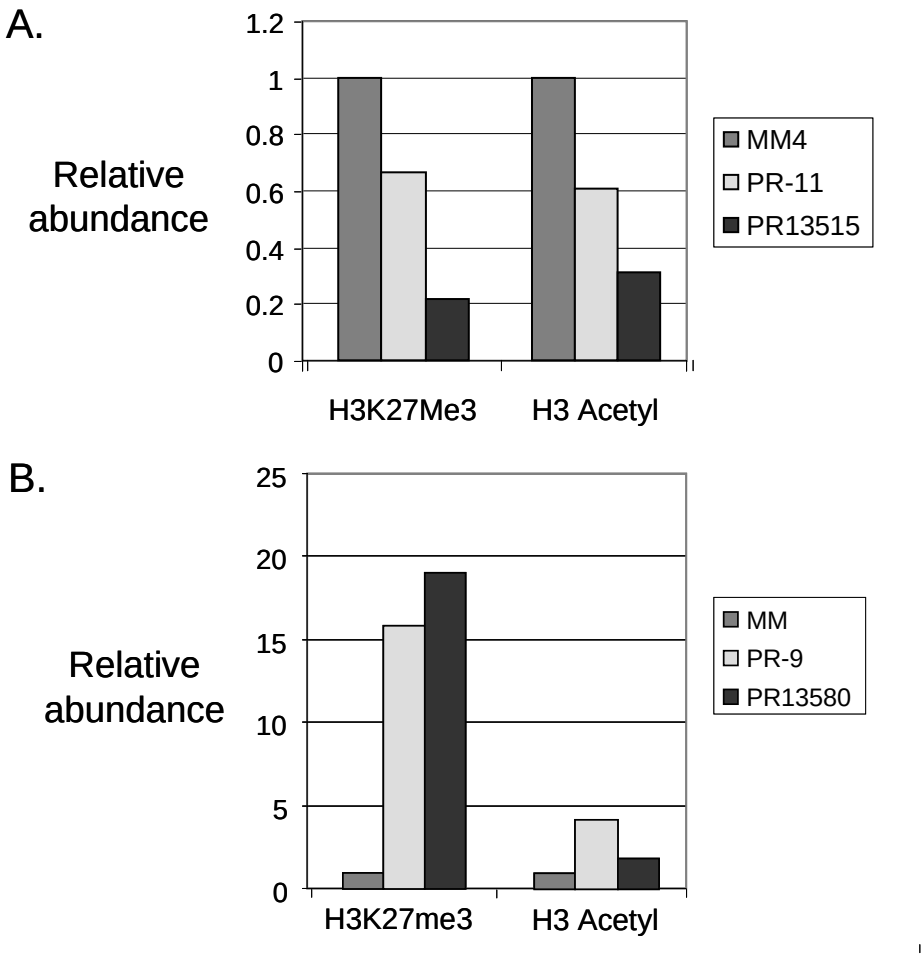


Figure 56: Histone acetylation and methylation. (A) In MCF7 cells, activating agRNAs PR-11 and PR13515 reduce H3K27 trimethylation and histone H3 acetylation. (B) In T47D cells, silencing agRNAs PR-9 and PR13580 increases H3K27 trimethylation and histone H3 acetylation.

Reviewers for the Yue et al paper covering Chapter 5 asked about involvement of other suggested members of the RISC complex with agRNAs. These proteins were PACT, Dicer, RHA, and TRBP. Dicer and TRBP have been suspected members of a RISC complex due to their homologues being involved in *in vitro* reconstituted RISC from *D. melanogaster (343, 344)* and, more recently, due to electron microscopy based structures of *Giardhia* RISC (*345*). Biochemical purification of tagged antibodies suggest that these human proteins form a complex (*346*). Commercially available TRBP antibodies are not recommended for immunoprecipitation experiments. However, I used a commercially available anti-Dicer antibody for immunoprecipitation. This antibody is recommended for immunoprecipitation and S. Younger has successfully used it to perform westerns of Dicer (S. Younger unpublished results).

I performed RNA immunoprecipitation using anti-Dicer antibody (n = 2) for both 5' and 3' targeting silencing agRNAs. Counterintuitively, Dicer pulled down the antisense transcript AT-2 in MM treated samples but not in agRNA treated samples. Y. Chu performed a ChIP experiment on the PR promoter and the 3' terminus of PR using this anti-Dicer antibody (n = 1) for 5' and 3' targeting activating agRNAs in MCF7 cells and also found Dicer associated with the PR promoter and 3' terminus in MM treated but not in agRNA treated samples. If this result reflects real reorganization of Dicer protein on the PR promoter, it is difficult to imagine the reason for this observation. One interpretation would be that this is simply pulldown of a nonspecific protein. However, it would be helpful to know if Dicer protein can be detected by western in the nucleus. In section 7.3, I show that knockdown of Dicer protein results in significant upregulation of PR mRNA in MCF7 cells. This result may simply be a result of the loss of miRNAs that regulate a transcriptional network regulating PR or that regulate PR mRNA itself.

The protein PACT is similar in structure to TRBP and flag-tagged protein biochemically purifies with human RISC complex (*347*). However, a confounding issue with biochemical purification of RNA binding proteins is that their co-purification may only reflect that they both bind the same target RNA transcript. Also, loss of PACT seems to disrupt the formation of mature miRNA. I performed RNA immunoprecipitation using an anti-PACT antibody for both 5' and 3' targeting activating and silencing agRNAs

in MCF7 and T47D cells and did not recover either the 5' or 3' associated noncoding RNA transcripts. This may be due to the quality of the antibody or that PACT is simply not in the nucleus.

Finally, the protein RHA is a member of the RNAP2 complex. A 2007 paper reports that RHA may be a member of the RISC complex and somehow help RISC cleave RNA targets (*348*). As mentioned above, co-purification of two RNA binding proteins is difficult to interpret. Also, knockdown of RHA slightly reduces siRNA efficiency *in vivo* but this could also be because of offtarget effects on gene expression since RHA is an important member of the RNAP2 complex. Finally, *in vitro* RHA enhances argonaute cleavage of an RNA target. This may be simply due to the fact that RHA, an RNA helicase, binds and loosens the secondary structure of an RNA target allowing better access for argonaute proteins. I ordered the same RHA antibody (Vaxron) used in the 2007 Robb and Rana paper to perform RIP. The product insert indicated several bands detected by western raising concerns about this antibody's specificity. I performed RIP using this antibody in MCF7 or T47D cells treated with either 5' or 3' targeting activating or silencing agRNAs. RHA immunoprecipitation recovered the 5' and 3' transcripts in every sample. This may be mediated through RHA's interactions with RNAP2 or simply due to nonspecificity of the antibody.

Previous reports have suggested an association between argonaute proteins and Gemin proteins, Gemin3 and Gemin4 (*349, 350*). For this reason I used an antibody to Gemin3 (clone 12H12, Sigma Aldrich) to perform ChIP in T47D (n = 3) and MCF7 (n = 1) cells treated with either MM, silencing, or activating agRNA. For T47D, a slight increase in Gemin3 is seen associated with the PR promoter in PR-9 treated cells (**Fig 55d**). For MCF7, Gemin3 did not pull down over negative IgG control (**Fig 55f**).

7.3 CHIP-SEQUENCING OF ARGONAUTE

Multiple lines of evidence is developing to indicate that argonaute is present in the nucleus of human cells. A question remains as to what endogenous argonaute does in the nucleus. Argonaute may be recruited to gene promoters across the genome to activate

and silence transcription. Argonaute may be recruited to heterochromatic regions as part of a human RITS complex. Argonaute may also be recruited to intergenic regions to regulate expression of noncoding RNAs. To answer this question I set out to perform a ChIP experiment to recover argonaute proteins and the perform high-throughput Selexa sequencing to sequence ever DNA sequence associated with argonaute proteins. I planned to have the company GenPathway Inc perform the ChIP and sequencing. For a positive control, I planned to use PR-11 recruitment of argonaute to the PR promoter. To test the biological relevance of detected argonaute-DNA interactions, I planned to deplete cells of small RNAs by knockdown of the protein Dicer. I transfected 10 million cells with either MM, PR-11, or siRNA targeting Dicer (siDicer). Both PR-11 and siDICER caused pronounced upregulation of PR expression.

To test the efficiency of argonaute pulldown I performed ChIP of argonaute for MM, PR-11, and siDicer treated cells. Here I encountered technical difficulty in that some samples pulled down the PR promoter efficiently and many did not. After several attempts, I did not feel that the ChIP experiment using the Mourelatos antibody was robust enough to warrant proceeding to Solexa sequencing. I also tried to use an alternative ChIP protocol used by D. Hardy for Schwartz et al (*351*), which uses three times more lysate and uses one tenth the concentration of SDS in the lysis buffer. After sonication, the lysis buffer remained cloudy, which I believe is due to chromatin precipitating out of solution. I did not recover any ChIP signal using this protocol.

Another approach to this experiment would be to perform RNA immunoprecipitation from the nuclear fraction followed by sequencing of all RNAs recovered with argonaute proteins. Y. Chu has also worked extensively with argonaute RIPs and this protocol is much more robust than argonaute ChIP experiments. Using the the new ABI Solid3 system, the entire transcriptome associated with argonaute proteins could be sequenced, which would give analogous information to ChIP-sequencing. A Dicer knockdown would test if argonaute proteins are recruited in a Dicer dependent fashion. Finally, ChIP-sequencing of RNAP2 for MM or siDicer treated samples would allow one to test if changes in argonaute association with nearby encoded RNAs correlate to changes in transcription.

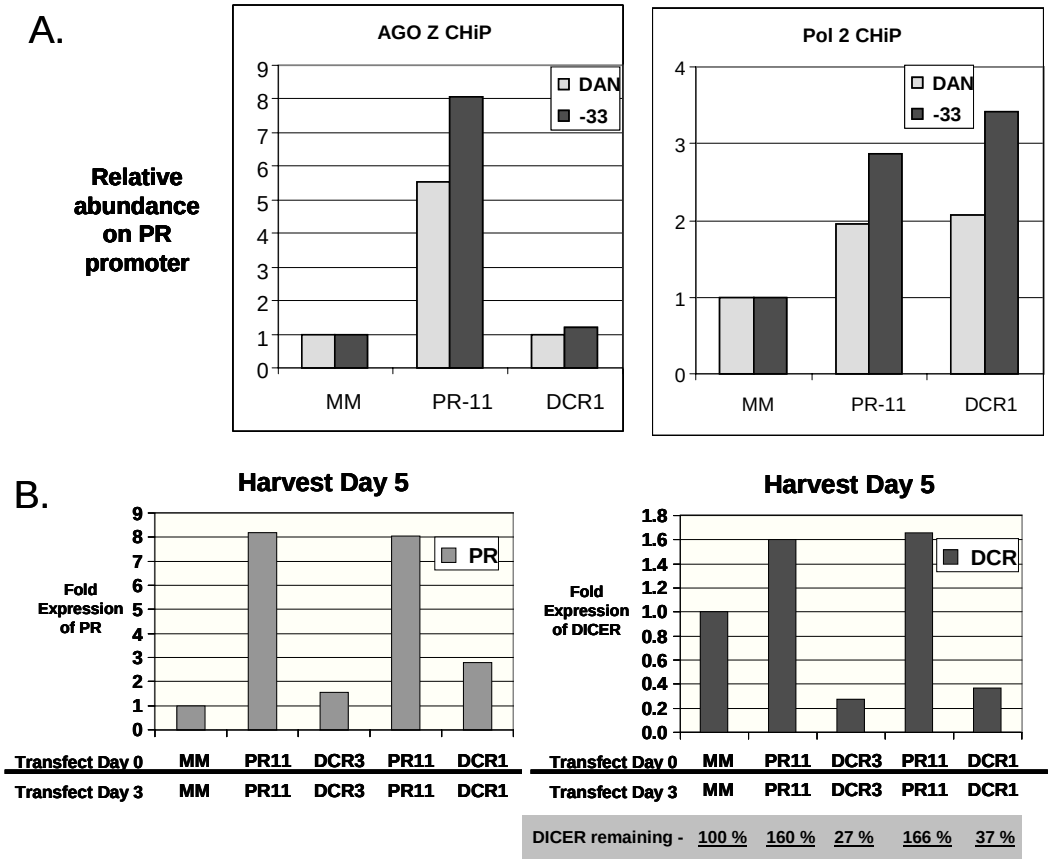


Figure 57: Effects of Dicer knockdown of PR. (A) Activating agRNA PR-11 in MCF7 cells induced recruitment of argonaute proteins detected by ChIP, although this assay struggled to reproduce consistently. However, PR-11 and the siRNA DCR1 targeting Dicer mRNA both upregulated transcription of PR detected by ChIP of RNAP2 at the promoter of PR. (B) Using a double transfection protocol, the siRNA DCR3 caused less upregulation of PR than DCR1. Both DCR3 and DCR1 appear to knockdown Dicer mRNA efficiently.

Additionally, an attractive approach for discovering endogenous agRNAs would be to sequence small RNAs recovered with argonaute proteins from the nuclear fraction. Numerous sequences have been deposited in miRBase (http://www.mirbase.org/) from sequencing of small RNAs associated with argonaute proteins recovered from whole cell fractions. However, there is no way to know that any of these sequences are actually transported into the nucleus, a requirement for them to be able to act as an endogenous agRNA. Since the role of argonaute in the nucleus remains unknown, sequencing of RNAs that are transported into the nucleus with argonaute would be a significant advance in our understanding of this protein's role.

7.4 CO-IMMUNOPRECIPITATION OF ARGONAUTE WITH RNAP2

Kim et al. reported that Ago1 interacts directly with RNAP2 (*256*). However, our current data suggests that while Ago1 may interact with agRNAs, it is not required for agRNA activity (Y. Chu unpublished). I decided to attempt a co-immunoprecipitation experiment to test if an argonaute interacts with RNAP2 directly. I used the Mourelatos anti-argonaute that pulls down all four argonaute proteins. I also pulled down RNAP2 using an anti-RNAP2 antibody.

I treated either whole cell lysate or nuclear fraction with RNase A or no treatment and incubated lysates with 4 μg of anti-argonaute, anti-RNAP2, or negative IgG antibody. I washed the beads four times with 1 mL of lysis buffer. I eluted protein from beads using elution buffer. I then performed a western using elute on a 7% polyacrylamide gel. I probed the blot using antibodies to either argonaute, RNAP2, Ago1, Ago2, or FXR1. The Mourelatos anti-argonaute antibody recovered Ago2 but did not seem to pull down Ago1 from the nuclear fraction. There was no observed pulldown of RNAP2 with argonaute. The RNAP2 antibody recovered RNAP2 but not any argonaute protein.

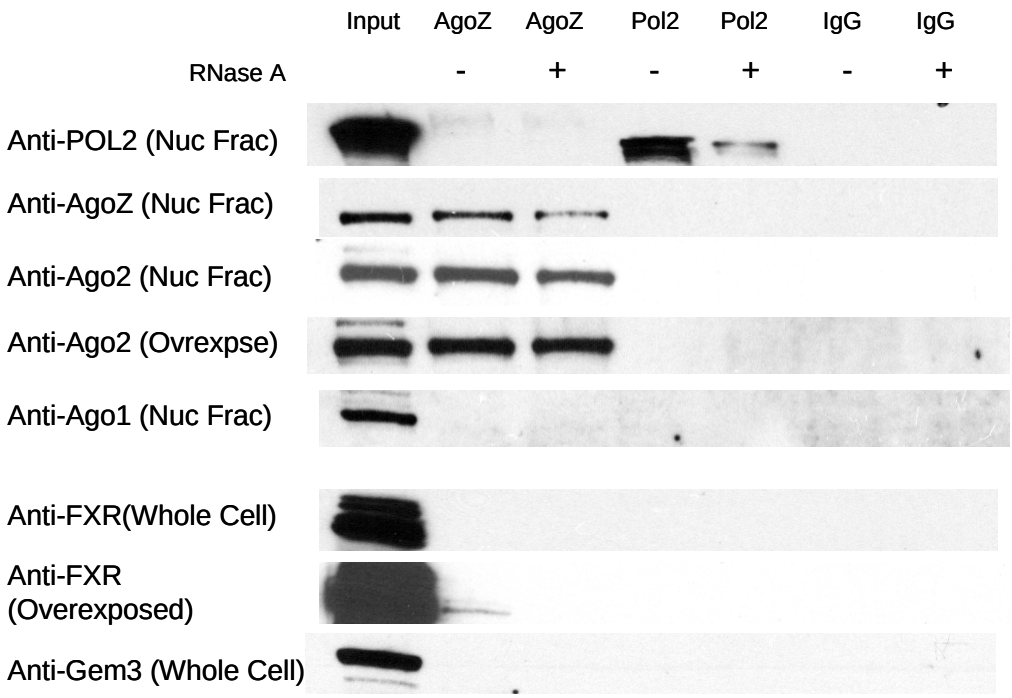


Figure 58: Co-immunoprecipitation of argonaute and RNAP2. I performed a immunoprecipitation with antibodies against argonaute and RNAP2. Samples were either treated with RNase A or untreated. For pulldown from the nuclear fraction I performed westerns using antibodies against RNAP2, argonaute, Ago2, and Ago1. For pulldown from the whole cell fraction, I performed westerns using antibodies against FXR1 and Gemin3.

Argonaute also did not pulldown FXR1. Overexposing the FXR1 western revealed a faint band in the argonaute pulldown lane but this may be just lane bleed from the input lane. This result is inconclusive since there is no positive control such as an interaction between FXR1 or Gemin3 and Ago. One interpretation could be that the lysis buffer is too strong and disrupts protein complexes. It would be interesting to get a positive control working to allow purification of nuclear argonaute complexes and investigate its protein interacting partners – such as potentially the HP1 or polycomb proteins. If a good purification of nuclear argonaute complexes is achieved, one could also take a more unbiased approach to co-factor discovery using a technique such as mass spectroscopy.

7.5 THE ROLE OF CHROMATIN LOOPS IN TRANSCRIPTION REGULATION

Chapter 6 investigated the role of chromatin loops in agRNA activity. While no changes in chromatin loops was observed, the existence of chromatin loops supports a model that 3' targeting agRNAs can regulate the gene promoter due to its physical close proximity. While changes in looping between the 5' and 3' ends of the PR gene does not correlate with changes in transcription states, there remains the possibility that larger remodeling of chromatin structure may play a role in this regulation.

Recent studies suggest that transcriptionally active and inactive chromatin is sequestered to different sublocales within the nucleus. Transcriptionally active centers within the nucleus are hypothesized to be near nuclear speckles, which are enriched for splicing machinery (*352*). It is possible that while local chromatin structure, such as the interaction between the 5' and 3' ends of the PR gene remain intact as transcription is activated or silenced by agRNAs, gross chromatin structure such as the association of the PR locus with nuclear speckles may change (*340*). The prediction would be that a particular gene may be packaged with other genes that have a similar transcriptional state, whether it is high transcription, low transcription, or no transcription.

To test this hypothesis would require multiple approaches. First, a more genome wide 3C approach to test chromatin structure changes can be taken. For this approach, 3C

samples are prepared as described in chapter 6 but detection is by high throughput sequencing rather than PCR. Interactions can be validated by PCR. Another line of validation would be to perform FISH in cell culture to visualize the co-localization of the PR locus with either other loci predicted to interact by 3C or with nuclear speckles. A final line of validation would be to use ChIP to measure changes in association of the PR locus DNA with structural proteins that uniquely compose nuclear speckles, the nucleolus, or the nuclear lamin.

CHAPTER EIGHT: Conclusions and future directions

8.1 CONCLUSIONS

What can be understood from my research is antigene RNA mediates their regulation of transcript through interactions with noncoding RNAs. This conclusion is supported by the following observations: 1) antisense gapmers knocking down the noncoding RNA in the PR promoter reverses gene activation, 2) biotinylated antigene RNAs pulldown noncoding RNAs and not genomic DNA (S. Younger), and 3) argonaute proteins are recruited to interact with noncoding RNAs by antigene RNAs.

This conclusion is meaningful in the development of agRNAs as a potential therapeutic because it identifies the target molecule that agRNAs bind. RNA as a drug target has many important structural and chemical differences from double stranded chromosomal DNA. This will help in the design of better antigene RNAs to both silence and activate gene expression.

The broader implications of knowing this target remain to be seen. Questions remain unanswered such as can noncoding RNAs that are too lowly abundant to be detected by RACE and cloned be targeted – such as intron sequences or other undetected sense or antisense RNA species. Also, what is the nature of the noncoding RNA's interaction to the chromatin and how does that effect agRNA activity? Finally, is it the nature of the interaction of the noncoding RNA with the agRNA that determines whether gene activation or silencing is observed, or perhaps the identity of the proteins involved in the interaction?

After this conclusion, my work characterizing regulation at the 3' end of genes suggests that interactions of agRNAs with noncoding RNAs can regulate transcription from a distance. I hypothesize that the existence of long distance interactions between noncoding RNAs and chromosomal DNA allows argonaute proteins recruited a site hundreds of thousands of nucleotides away from a gene promoter to physically interact with noncoding RNAs at the gene promoter and thereby regulate transcription. This

conclusion is supported by the observations that 1) agRNAs recruit argonaute proteins to interact with noncoding RNAs at both the 5' and 3' ends of the PR gene regardless of which end is targeted, and 2) long distance interactions between chromosomal DNA can be detected between the 5' and 3' ends of PR and BRCA1 in both treated and untreated MCF7 and T47D cells.

The broader implications of this action at a distance remain to be seen. Chromatin loops other than these observed between the 5' and 3' ends of genes may also serve as substrates for agRNA regulation at a distance. In addition to providing additional targets through which to regulate transcription, this mechanism may also allow agRNAs to regulate transcription of adjacent genes upstream and downstream of the targeted gene. Genes such as progesterone receptor lie in relatively gene poor regions of the genome with no other protein-coding genes lying within 100 kb of the 5' or 3' ends of the PR gene. However, other more gene rich regions of the genome may provide a better model for future testing of this action at a distance model.

8.2 FUTURE DIRECTIONS

8.2.1 The role of argonaute in the nucleus

Obvious biological experiments such as the biochemical purification of an argonaute-agRNA complex with its protein cofactors are beyond the scope of what is the core focus of the Corey lab. A biological question concerning the possibility of naturally occurring agRNAs is a goal that may fall within the scope of the Corey lab's core focus. In order to ensure success of the search for endogenous agRNAs, I think that the question should be broken into smaller key advances, the accomplishment of which may prove a significant advancement in and of themselves.

The establishment of argonaute proteins being present in the nucleus is only now becoming appreciated in mammals. Thus, any experiment investigating the role of argonaute in the nucleus has a high potential to have profound results. Experiments may include RIP-sequencing of small RNAs associated with argonaute isolated from nuclei.

Also, RIP-sequencing of RNA transcripts associated with argonaute isolated from nuclei may lead to insight about the role of argonaute in the nucleus. Finally, a knockdown of argonaute or dicer proteins and a microarray analysis of RNA expression in the nucleus or even a ChIP-chip experiment of RNA polymerase II to see genome wide changes in transcription may also present leads concerning the role of argonaute in the nucleus.

Isolating nuclei focuses the investigation away from standard RNAi, which occurs in the cell cytoplasm. However, this isolation does not preclude that observations in the nucleus are not influenced by standard RNAi in the cytoplasm. Therefore, to promote the greatest possibility of uncovering a role for argonaute proteins in the nucleus, the investigation of all three above mentioned experiments in parallel would allow each technique to reinforce results gotten by other techniques and provide the best screen for leads that should be analyzed more closely. Even though a small RNA may be present in the nucleus, lowly expressed small RNAs may prove technically difficult to perform follow up investigations on leading to uninterpretable results. Thus RIP-sequencing of small RNAs would provide insight into which small RNA would be best to investigate in a particular cell line of interest. Furthermore, any small RNA sequence can be predicted to target thousands, if not tens of thousands, of sequences in the human genome. Thus knowing which interactions actually occur and in high enough abundance to be detectable would further guide follow up investigations. Finally, small RNAs may recognize a noncoding RNA in the nucleus with no observable biological outcome. These may be technical artifacts or simply due to the complexity of inputs into a biological signaling pathway. Therefore, knowing which interactions give the strongest observable biological outcome will make follow up investigations much simpler.

There is no guarantee that argonaute's primary role in the nucleus is to be involved in agRNA related activity. Thus an unbiased investigation of argonaute's role in the nucleus may produce any number of activities; of which, agRNA related gene silencing and activation activity may only represent a small fraction. However, the discovery of any role for argonaute in the nucleus would be a major breakthrough in RNAi biology regardless of whether or not it closely resembles the activity seen for agRNAs. Furthermore, endogenously expressed agRNAs may have important

mechanistic differences from exogenously transfected agRNAs that would preclude their discovery by a more narrowly focused investigation.

8.2.2 The role of chromatin modifications in agRNA activity

A simple experiment that has already begun to be explored in the Corey lab is the role of chromatin modification in agRNA activity. However, it is difficult to design experiments for this investigation with any high promise of producing mechanistic insight. The problems include questions about the biological relevance of observed chromatin modifications. It would not be a significant advance in the understanding of agRNAs to merely catalogue a list of chromatin modifications that loosely to moderately associate with agRNA activity. However, a strongly and highly reproducible chromatin modification associated with either gene silencing or activation may be a crucial clue granting insight into the mechanism of agRNAs. On the other hand, the proteins responsible for some chromatin modifications have only been characterized for certain chromatin modifications under certain circumstances in human cells. Thus the observation of any chromatin modification highly correlated to agRNA activity may not produce any actionable hypotheses.

The most insight concerning the role of chromatin modifications in agRNA activity would be if these investigations were correlated with a study biochemically purifying the protein cofactors associated with argonaute in the nucleus. In this manner, even if poorly understood chromatin modifications were found to highly correlate with agRNA activity, knowing protein cofactors involved may reveal the role that these modifications play in the mechanism. If relatively well understood chromatin modifications are found to closely associate with agRNA activity, there is no guarantee that their mechanistic role in agRNA activity will be the same as observed for transcription regulation by other transcription factor and chromatin remodeling complexes.

8.2.3 The biological role of chromatin looping

Chromatin looping is potentially profoundly important for regulation of transcription by agRNAs. However, the mechanism of chromatin looping and its role in normal biology is poorly understood. For this reason, any investigation into the role of chromatin looping may potentially have far reaching impact beyond the short term goal of understanding the mechanism of agRNAs.

On the basis of my research, changes in chromatin looping do not appear to be the mechanism by which agRNA regulate transcription. However, the model that gene loops are required for agRNAs to regulate from a distance remains to be rigorously tested. What would be desireable is if a chromatin loop could be tested that is known to be reversible by some treatment, such as estradiol. For this, a well characterized, long-distance enhancer interaction may suit this study better than the hope that some 5' to 3' end interaction will be found to be reversible under some condition. If it could be shown that a chromatin loop is reversible, and agRNAs targeting the distant site can be found to regulate transcription, then the necessity of gene loops for agRNA action at a distance could be rigorously tested.

If gene loops are required for agRNA action at a distance, then the conditions under which these interactions occur could have a profound impact on the potential for agRNAs targeting distant sites to be used as a drug therapeutic. Simple questions such as whether gene looping changes throughout cell cycle can have important implications. If gene loops only occur during a particular step in cell cycle, than the existence of a particular loop and possibly the ability of agRNAs to regulate at a distance may depend on whether or not cells are dividing. This may impact the specificity of agRNA therapeutics designed to target rapidly dividing cancer cells. Furthermore, do gene loops differ from tissue to tissue? The answer to this question may allow a mechanism to design tissue specific agRNAs.

Beyond the development of agRNAs as a therapeutic, the answer these questions may have a broader impact on cell biology. Questions about how chromatin is packed in the nucleus and subnuclear organization form a new frontier in cell biology. How robust

is chromatin looping? What is the probability that two cells of the same tissue would have the same chromatin loop and, by extrapolation, the same global chromatin structure? To answer these questions, chromatin conformation capture alone cannot be expected to be a robust enough technique. Instead the combination of 3C with other techniques such as RIP, microscopy, and RNA-TRAP (*179*), may provide a more robust platform for investigating these questions.

Investigating chromatin loops under the context of agRNA mechanisms also presents an advantage to the study of chromatin loops in general. agRNAs allow for the sequence specific recognition of a noncoding RNA by argonaute proteins. This allows for the use of RIP experiments to validate 3C results. Also, agRNAs recruit argonaute proteins resulting in a biological output, either transcriptional silencing or activation. If this biological output depends on chromatin looping, using agRNAs may allow for the development of a screen to test for conditions under which chromatin looping occurs.

8.2.4 The biological role of noncoding RNAs

Like chromatin loops, the biological significance of noncoding RNAs remains to be seen. The answer to this question is important to the development of agRNAs as a therapeutic. Presumably, a noncoding RNA such as AT-2 has a biological function beyond merely serving as a substrate for agRNAs. Uncovering the biological function of these noncoding RNAs may unveil a broader output from agRNAs, beyond merely affecting transcription of the targeted protein-coding gene and its downstream signaling cascade. As seen for miRNAs, the output from affecting expression of a noncoding RNA such as AT-2 may be quite broad.

BIBLIOGRAPHY

1. A. Z. Fire, *Cell Death Differ* **14**, 1998 (Dec, 2007).

2. C. C. Mello, *Cell Death Differ* **14**, 2013 (Dec, 2007).

3. R. C. Lee, R. L. Feinbaum, V. Ambros, *Cell* **75**, 843 (Dec 3, 1993).

4. M. A. Matzke, A. Matzke, *Plant Physiol* **107**, 679 (Mar, 1995).

5. A. Fire *et al.*, *Nature* **391**, 806 (Feb 19, 1998).

6. H. Tabara *et al.*, *Cell* **99**, 123 (Oct 15, 1999).

7. J. B. Morel *et al.*, *Plant Cell* **14**, 629 (Mar, 2002).

8. S. M. Hammond, S. Boettcher, A. A. Caudy, R. Kobayashi, G. J. Hannon, *Science* **293**, 1146 (Aug 10, 2001).

9. A. J. Hamilton, D. C. Baulcombe, *Science* **286**, 950 (Oct 29, 1999).

10. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (Mar 31, 2000).

11. F. Wianny, M. Zernicka-Goetz, *Nat Cell Biol* **2**, 70 (Feb, 2000).

12. S. M. Elbashir *et al.*, *Nature* **411**, 494 (May 24, 2001).

13. J. Martinez, A. Patkaniowska, H. Urlaub, R. Luhrmann, T. Tuschl, *Cell* **110**, 563 (Sep 6, 2002).

14. D. R. Corey, *Nat Chem Biol* **3**, 8 (Jan, 2007).

15. M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, *Science* **294**, 853 (Oct 26, 2001).

16. R. C. Lee, V. Ambros, *Science* **294**, 862 (Oct 26, 2001).

17. N. C. Lau, L. P. Lim, E. G. Weinstein, D. P. Bartel, *Science* **294**, 858 (Oct 26, 2001).

18. M. Ghildiyal, P. D. Zamore, *Nat Rev Genet* **10**, 94 (Feb, 2009).

19. M. A. Matzke, M. Primig, J. Trnovsky, A. J. Matzke, *EMBO J* **8**, 643 (Mar, 1989).

20. M. A. Matzke, A. J. Matzke, *Plant Mol Biol* **16**, 821 (May, 1991).

21. M. F. Mette, J. van der Winden, M. A. Matzke, A. J. Matzke, *EMBO J* **18**, 241 (Jan 4, 1999).

22. Y. D. Park *et al.*, *Plant J* **9**, 183 (Feb, 1996).

23. M. Wassenegger, S. Heimes, L. Riedel, H. L. Sanger, *Cell* **76**, 567 (Feb 11, 1994).

24. M. F. Mette, W. Aufsatz, J. van der Winden, M. A. Matzke, A. J. Matzke, *EMBO J* **19**, 5194 (Oct 2, 2000).

25. B. Huettel *et al.*, *Biochim Biophys Acta* **1769**, 358 (May-Jun, 2007).

26. M. Matzke, T. Kanno, L. Daxinger, B. Huettel, A. J. Matzke, *Curr Opin Cell Biol* **21**, 367 (Jun, 2009).

27. A. Hamilton, O. Voinnet, L. Chappell, D. Baulcombe, *EMBO J* **21**, 4671 (Sep 2, 2002).

28. S. W. Chan, I. R. Henderson, S. E. Jacobsen, *Nat Rev Genet* **6**, 351 (May, 2005).

29. S. W. Chan *et al.*, *Science* **303**, 1336 (Feb 27, 2004).

30. Y. Qi *et al.*, *Nature* **443**, 1008 (Oct 26, 2006).

31. A. J. Herr, M. B. Jensen, T. Dalmay, D. C. Baulcombe, *Science* **308**, 118 (Apr 1, 2005).

32. Y. Onodera *et al.*, *Cell* **120**, 613 (Mar 11, 2005).

33. D. Pontier *et al.*, *Genes Dev* **19**, 2030 (Sep 1, 2005).

34. T. S. Ream *et al.*, *Mol Cell* **33**, 192 (Jan 30, 2009).

35. A. T. Wierzbicki, T. S. Ream, J. R. Haag, C. S. Pikaard, *Nat Genet* **41**, 630 (May, 2009).

36. A. T. Wierzbicki, J. R. Haag, C. S. Pikaard, *Cell* **135**, 635 (Nov 14, 2008).

37. B. Huettel *et al.*, *EMBO J* **25**, 2828 (Jun 21, 2006).

38. L. Aravind, H. Watanabe, D. J. Lipman, E. V. Koonin, *Proc Natl Acad Sci U S A* **97**, 11319 (Oct 10, 2000).

39. T. A. Volpe *et al.*, *Science* **297**, 1833 (Sep 13, 2002).

40. I. A. Drinnenberg *et al.*, *Science*, (Sep 10, 2009).

41. A. Verdel *et al.*, *Science* **303**, 672 (Jan 30, 2004).

42. K. Noma *et al.*, *Nat Genet* **36**, 1174 (Nov, 2004).

43. M. R. Motamedi *et al.*, *Cell* **119**, 789 (Dec 17, 2004).

44. M. Buhler, A. Verdel, D. Moazed, *Cell* **125**, 873 (Jun 2, 2006).

45. M. Buhler, *Chromosoma* **118**, 141 (Apr, 2009).

46. S. I. Grewal, S. Jia, *Nat Rev Genet* **8**, 35 (Jan, 2007).

47. E. Nicolas *et al.*, *Nat Struct Mol Biol* **14**, 372 (May, 2007).

48. T. Iida, J. Nakayama, D. Moazed, *Mol Cell* **31**, 178 (Jul 25, 2008).

49. K. L. Huisinga, S. C. Elgin, *Biochim Biophys Acta* **1789**, 3 (Jan, 2009).

50. A. A. Aravin *et al.*, *Curr Biol* **11**, 1017 (Jul 10, 2001).

51. A. Schmidt *et al.*, *Genetics* **151**, 749 (Feb, 1999).

52. D. N. Cox *et al.*, *Genes Dev* **12**, 3715 (Dec 1, 1998).

53. D. N. Cox, A. Chao, H. Lin, *Development* **127**, 503 (Feb, 2000).

54. C. D. Malone, G. J. Hannon, *Cell* **136**, 656 (Feb 20, 2009).

55. E. Sarot, G. Payen-Groschene, A. Bucheton, A. Pelisson, *Genetics* **166**, 1313 (Mar, 2004).

56. M. Savitsky, D. Kwon, P. Georgiev, A. Kalmykova, V. Gvozdev, *Genes Dev* **20**, 345 (Feb 1, 2006).

57. D. Reiss, T. Josse, D. Anxolabehere, S. Ronsseray, *Mol Genet Genomics* **272**, 336 (Oct, 2004).

58. A. A. Aravin *et al.*, *Dev Cell* **5**, 337 (Aug, 2003).

59. V. V. Vagin *et al.*, *Science* **313**, 320 (Jul 21, 2006).

60. K. Saito *et al.*, *Genes Dev* **20**, 2214 (Aug 15, 2006).

61. W. Deng, H. Lin, *Dev Cell* **2**, 819 (Jun, 2002).

62. A. Aravin *et al.*, *Nature* **442**, 203 (Jul 13, 2006).

63. A. Girard, R. Sachidanandam, G. J. Hannon, M. A. Carmell, *Nature* **442**, 199 (Jul 13, 2006).

64. N. C. Lau *et al.*, *Science* **313**, 363 (Jul 21, 2006).

65. J. Brennecke *et al.*, *Cell* **128**, 1089 (Mar 23, 2007).

66. L. S. Gunawardane *et al.*, *Science* **315**, 1587 (Mar 16, 2007).

67. A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, G. J. Hannon, *Science* **316**, 744 (May 4, 2007).

68. A. A. Aravin *et al.*, *Mol Cell* **31**, 785 (Sep 26, 2008).

69. K. V. Morris, S. W. Chan, S. E. Jacobsen, D. J. Looney, *Science* **305**, 1289 (Aug 27, 2004).

70. G. J. Hannon, J. J. Rossi, *Nature* **431**, 371 (Sep 16, 2004).

71. A. L. Jackson *et al.*, *Nat Biotechnol* **21**, 635 (Jun, 2003).

72. S. P. Persengiev, X. Zhu, M. R. Green, *RNA* **10**, 12 (Jan, 2004).

73. C. A. Sledz, M. Holko, M. J. de Veer, R. H. Silverman, B. R. Williams, *Nat Cell Biol* **5**, 834 (Sep, 2003).

74. H. Kawasaki, K. Taira, *Nature* **431**, 211 (Sep 9, 2004).

75. D. Castanotto *et al.*, *Mol Ther* **12**, 179 (Jul, 2005).

76. P. J. Paddison *et al.*, *Nat Methods* **1**, 163 (Nov, 2004).

77. P. J. Paddison, A. A. Caudy, R. Sachidanandam, G. J. Hannon, *Methods Mol Biol* **265**, 85 (2004).

78. K. Suzuki *et al.*, *j RNAi Gene Silencing* **1**, 66 (2005, 2005).

79. E. P. Murchison, J. F. Partridge, O. H. Tam, S. Cheloufi, G. J. Hannon, *Proc Natl Acad Sci U S A* **102**, 12135 (Aug 23, 2005).

80. R. Benetti *et al.*, *Nat Struct Mol Biol* **15**, 268 (Mar, 2008).

81. L. Sinkkonen *et al.*, *Nat Struct Mol Biol* **15**, 259 (Mar, 2008).

82. S. Vasudevan, J. A. Steitz, *Cell* **128**, 1105 (Mar 23, 2007).

83. S. Vasudevan, Y. Tong, J. A. Steitz, *Science* **318**, 1931 (Dec 21, 2007).

84. S. Vasudevan, Y. Tong, J. A. Steitz, *Cell Cycle* **7**, 1545 (Jun 1, 2008).

85. F. Crick, *Nature* **227**, 561 (Aug 8, 1970).

86. F. H. C. Crick, *Symposia of the Society for Experimental Biology: The Biological Replication of Macromolecules* **12**, 152 (1958).

87. D. Baltimore, *Nature* **226**, 1209 (Jun 27, 1970).

88. H. M. Temin, S. Mizutani, *Nature* **226**, 1211 (Jun 27, 1970).

89. T. R. Cech, A. J. Zaug, P. J. Grabowski, *Cell* **27**, 487 (Dec, 1981).

90. A. L. Bothwell, B. C. Stark, S. Altman, *Proc Natl Acad Sci U S A* **73**, 1912 (Jun, 1976).

91. F. H. Crick, *J Mol Biol* **38**, 367 (Dec, 1968).

92. H. B. White, 3rd, *J Mol Evol* **7**, 101 (Mar 29, 1976).

93. T. R. Cech, *Cell* **136**, 599 (Feb 20, 2009).

94. B. Alberts *et al.*, *Molecular biology of the cell*. (Garland Science, New York, ed. 4, 2002).

95. A. Tissieres, J. D. Watson, *Nature* **182**, 778 (Sep 20, 1958).

96. J. W. Littlefield, E. B. Keller, J. Gross, P. C. Zamecnik, *J Biol Chem* **217**, 111 (Nov, 1955).

97. P. B. Moore, T. A. Steitz, *Annu Rev Biochem* **72**, 813 (2003).

98. M. Takanami, *Biochim Biophys Acta* **55**, 132 (Jan 22, 1962).

99. M. B. Hoagland, M. L. Stephenson, J. F. Scott, L. I. Hecht, P. C. Zamecnik, *J Biol Chem* **231**, 241 (Mar, 1958).

100. H. Lodish *et al.*, *Molecular cell biology*. (W.H. Freeman and Co., New York, New York, ed. 5, 2004), pp. 973.

101. V. W. Yang, M. R. Lerner, J. A. Steitz, S. J. Flint, *Proc Natl Acad Sci U S A* **78**, 1371 (Mar, 1981).

102. A. A. Patel, J. A. Steitz, *Nat Rev Mol Cell Biol* **4**, 960 (Dec, 2003).

103. R. Reddy *et al.*, *J Biol Chem* **256**, 8452 (Aug 25, 1981).

104. G. Zieve, S. Penman, *Cell* **8**, 19 (May, 1976).

105. W. Filipowicz, V. Pogacic, *Curr Opin Cell Biol* **14**, 319 (Jun, 2002).

106. G. Dieci, M. Preti, B. Montanini, *Genomics* **94**, 83 (Aug, 2009).

107. S. Altman, *Mol Biosyst* **3**, 604 (Sep, 2007).

108. J. A. Goodrich, J. F. Kugel, *Crit Rev Biochem Mol Biol* **44**, 3 (Jan-Feb, 2009).

109. S. Xiao, F. Scott, C. A. Fierke, D. R. Engelke, *Annu Rev Biochem* **71**, 165 (2002).

110. R. W. Carthew, E. J. Sontheimer, *Cell* **136**, 642 (Feb 20, 2009).

111. T. A. Farazi, S. A. Juranek, T. Tuschl, *Development* **135**, 1201 (Apr, 2008).

112. D. P. Bartel, *Cell* **116**, 281 (Jan 23, 2004).

113. C. W. Greider, E. H. Blackburn, *Nature* **337**, 331 (Jan 26, 1989).

114. D. Shippen-Lentz, E. H. Blackburn, *Science* **247**, 546 (Feb 2, 1990).

115. D. P. Romero, E. H. Blackburn, *Cell* **67**, 343 (Oct 18, 1991).

116. J. Lingner *et al.*, *Science* **276**, 561 (Apr 25, 1997).

117. D. C. Zappulla, T. R. Cech, *Proc Natl Acad Sci U S A* **101**, 10024 (Jul 6, 2004).

118. F. Qiao, T. R. Cech, *Nat Struct Mol Biol* **15**, 634 (Jun, 2008).

119. S. R. Eddy, *Nat Rev Genet* **2**, 919 (Dec, 2001).

120. J. Vogel, *Mol Microbiol* **71**, 1 (Jan, 2009).

121. K. M. Wassarman, *Cell* **109**, 141 (Apr 19, 2002).

122. C. A. Brosnan, O. Voinnet, *Curr Opin Cell Biol* **21**, 416 (Jun, 2009).

123. D. D. Sledjeski, A. Gupta, S. Gottesman, *EMBO J* **15**, 3993 (Aug 1, 1996).

124. N. Majdalani, C. Cunning, D. Sledjeski, T. Elliott, S. Gottesman, *Proc Natl Acad Sci U S A* **95**, 12462 (Oct 13, 1998).

125. P. J. Mikulecky *et al.*, *Nat Struct Mol Biol* **11**, 1206 (Dec, 2004).

126. T. J. Soper, S. A. Woodson, *RNA* **14**, 1907 (Sep, 2008).

127. N. Majdalani, S. Chen, J. Murrow, K. St John, S. Gottesman, *Mol Microbiol* **39**, 1382 (Mar, 2001).

128. T. Updegrove, N. Wilf, X. Sun, R. M. Wartell, *Biochemistry* **47**, 11184 (Oct 28, 2008).

129. J. Hindley, *J Mol Biol* **30**, 125 (1967).

130. G. G. Brownlee, *Nat New Biol* **229**, 147 (1971).

131. K. M. Wassarman, G. Storz, *Cell* **101**, 613 (Jun 9, 2000).

132. J. E. Barrick, N. Sudarsan, Z. Weinberg, W. L. Ruzzo, R. R. Breaker, *RNA* **11**, 774 (May, 2005).

133. A. E. Trotochaud, K. M. Wassarman, *Nat Struct Mol Biol* **12**, 313 (Apr, 2005).

134. K. M. Wassarman, R. M. Saecker, *Science* **314**, 1601 (Dec 8, 2006).

135. E. Ullu, M. Melli, *Nucleic Acids Res* **10**, 2209 (Apr 10, 1982).

136. G. Storz, S. Altuvia, K. M. Wassarman, *Annu Rev Biochem* **74**, 199 (2005).

137. S. Murphy, A. Pierani, C. Scheidereit, M. Melli, R. G. Roeder, *Cell* **59**, 1071 (Dec 22, 1989).

138. V. T. Nguyen, T. Kiss, A. A. Michels, O. Bensaude, *Nature* **414**, 322 (Nov 15, 2001).

139. Z. Yang, Q. Zhu, K. Luo, Q. Zhou, *Nature* **414**, 317 (Nov 15, 2001).

140. A. A. Michels *et al.*, *EMBO J* **23**, 2608 (Jul 7, 2004).

141. S. Cho *et al.*, *EMBO J* **28**, 1407 (May 20, 2009).

142. T. Kuwabara, J. Hsieh, K. Nakashima, K. Taira, F. H. Gage, *Cell* **116**, 779 (Mar 19, 2004).

143. A. S. Krayev *et al.*, *Nucleic Acids Res* **10**, 7461 (Dec 11, 1982).

144. A. J. Fornace, Jr., J. B. Mitchell, *Nucleic Acids Res* **14**, 5793 (Jul 25, 1986).

145. T. A. Allen, S. Von Kaenel, J. A. Goodrich, J. F. Kugel, *Nat Struct Mol Biol* **11**, 816 (Sep, 2004).

146. C. A. Espinoza, T. A. Allen, A. R. Hieb, J. F. Kugel, J. A. Goodrich, *Nat Struct Mol Biol* **11**, 822 (Sep, 2004).

147. C. A. Espinoza, J. A. Goodrich, J. F. Kugel, *RNA* **13**, 583 (Apr, 2007).

148. R. B. Lanz *et al.*, *Cell* **97**, 17 (Apr 2, 1999).

149. M. Watanabe *et al.*, *EMBO J* **20**, 1341 (Mar 15, 2001).

150. H. Kawashima *et al.*, *Biochem J* **369**, 163 (Jan 1, 2003).

151. S. Chooniedass-Kothari *et al.*, *FEBS Lett* **566**, 43 (May 21, 2004).

152. R. B. Lanz, B. Razani, A. D. Goldberg, B. W. O'Malley, *Proc Natl Acad Sci U S A* **99**, 16081 (Dec 10, 2002).

153. T. M. DeChiara, E. J. Robertson, A. Efstratiadis, *Cell* **64**, 849 (Feb 22, 1991).

154. S. Banerjee, A. Smallwood, *Nat Genet* **11**, 237 (Nov, 1995).

155. G. Christofori, P. Naik, D. Hanahan, *Nat Genet* **10**, 196 (Jun, 1995).

156. K. L. Arney, *Trends Genet* **19**, 17 (Jan, 2003).

157. A. L. Webber, R. S. Ingram, J. M. Levorse, S. M. Tilghman, *Nature* **391**, 711 (Feb 12, 1998).

158. A. C. Bell, G. Felsenfeld, *Nature* **405**, 482 (May 25, 2000).

159. A. T. Hark *et al.*, *Nature* **405**, 486 (May 25, 2000).

160. W. Reik, A. Murrell, *Nature* **405**, 408 (May 25, 2000).

161. M. Constancia *et al.*, *Nat Genet* **26**, 203 (Oct, 2000).

162. A. Murrell, S. Heeson, W. Reik, *Nat Genet* **36**, 889 (Aug, 2004).

163. J. Q. Ling *et al.*, *Science* **312**, 269 (Apr 14, 2006).

164. E. D. Rubio *et al.*, *Proc Natl Acad Sci U S A* **105**, 8309 (Jun 17, 2008).

165. M. P. Lee *et al.*, *Proc Natl Acad Sci U S A* **96**, 5203 (Apr 27, 1999).

166. J. Y. Shin, G. V. Fitzpatrick, M. J. Higgins, *EMBO J* **27**, 168 (Jan 9, 2008).

167. N. J. Smilinich *et al.*, *Proc Natl Acad Sci U S A* **96**, 8064 (Jul 6, 1999).

168. R. R. Pandey *et al.*, *J Biol Chem* **279**, 52685 (Dec 10, 2004).

169. N. Thakur *et al.*, *Mol Cell Biol* **24**, 7855 (Sep, 2004).

170. D. Mancini-Dinardo, S. J. Steele, J. M. Levorse, R. S. Ingram, S. M. Tilghman, *Genes Dev* **20**, 1268 (May 15, 2006).

171. G. V. Fitzpatrick *et al.*, *Mol Cell Biol* **27**, 2636 (Apr, 2007).

172. R. R. Pandey *et al.*, *Mol Cell* **32**, 232 (Oct 24, 2008).

173. R. Terranova *et al.*, *Dev Cell* **15**, 668 (Nov, 2008).

174. R. Lyle *et al.*, *Nat Genet* **25**, 19 (May, 2000).

175. F. Sleutels, D. P. Barlow, *Mol Cell Biol* **21**, 5008 (Aug, 2001).

176. R. Zwart, F. Sleutels, A. Wutz, A. H. Schinkel, D. P. Barlow, *Genes Dev* **15**, 2361 (Sep 15, 2001).

177. F. Sleutels, R. Zwart, D. P. Barlow, *Nature* **415**, 810 (Feb 14, 2002).

178. C. I. Seidl, S. H. Stricker, D. P. Barlow, *EMBO J* **25**, 3565 (Aug 9, 2006).

179. T. Nagano *et al.*, *Science* **322**, 1717 (Dec 12, 2008).

180. C. J. Brown *et al.*, *Nature* **349**, 38 (Jan 3, 1991).

181. N. Brockdorff *et al.*, *Cell* **71**, 515 (Oct 30, 1992).

182. M. Zuccotti, M. Monk, *Nat Genet* **9**, 316 (Mar, 1995).

183. G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, *Nature* **379**, 131 (Jan 11, 1996).

184. J. T. Lee, R. Jaenisch, *Nature* **386**, 275 (Mar 20, 1997).

185. J. T. Lee, L. S. Davidow, D. Warshawsky, *Nat Genet* **21**, 400 (Apr, 1999).

186. J. T. Lee, N. Lu, *Cell* **99**, 47 (Oct 1, 1999).

187. J. T. Lee, *Cell* **103**, 17 (Sep 29, 2000).

188. Y. Ogawa, J. T. Lee, *Mol Cell* **11**, 731 (Mar, 2003).

189. W. Chao, K. D. Huynh, R. J. Spencer, L. S. Davidow, J. T. Lee, *Science* **295**, 345 (Jan 11, 2002).

190. C. L. Tsai, R. K. Rowntree, D. E. Cohen, J. T. Lee, *Dev Biol* **319**, 416 (Jul 15, 2008).

191. K. Plath *et al.*, *Science* **300**, 131 (Apr 4, 2003).

192. J. Silva *et al.*, *Dev Cell* **4**, 481 (Apr, 2003).

193. E. Bernstein *et al.*, *Mol Cell Biol* **26**, 2560 (Apr, 2006).

194. Y. Ogawa, B. K. Sun, J. T. Lee, *Science* **320**, 1336 (Jun 6, 2008).

195. J. L. Rinn *et al.*, *Cell* **129**, 1311 (Jun 29, 2007).

196. A. M. Khalil *et al.*, *Proc Natl Acad Sci U S A* **106**, 11667 (Jul 14, 2009).

197. X. Wang *et al.*, *Nature* **454**, 126 (Jul 3, 2008).

198. W. Yu *et al.*, *Nature* **451**, 202 (Jan 10, 2008).

199. I. Martianov, A. Ramadass, A. Serra Barros, N. Chow, A. Akoulitchev, *Nature* **445**, 666 (Feb 8, 2007).

200. S. Akoulitchev, T. P. Makela, R. A. Weinberg, D. Reinberg, *Nature* **377**, 557 (Oct 12, 1995).

201. S. W. Blume, Z. Meng, K. Shrestha, R. C. Snyder, P. D. Emanuel, *J Cell Biochem* **88**, 165 (Jan 1, 2003).

202. J. Wells, P. Held, S. Illenye, N. H. Heintz, *Mol Cell Biol* **16**, 634 (Feb, 1996).

203. J. Cheng *et al.*, *Science* **308**, 1149 (May 20, 2005).

204. P. Kapranov *et al.*, *Genome Res* **15**, 987 (Jul, 2005).

205. P. Kapranov *et al.*, *Science* **316**, 1484 (Jun 8, 2007).

206. P. Carninci *et al.*, *Science* **309**, 1559 (Sep 2, 2005).

207. S. Katayama *et al.*, *Science* **309**, 1564 (Sep 2, 2005).

208. M. Guttman *et al.*, *Nature* **458**, 223 (Mar 12, 2009).

209. M. C. Frith *et al.*, *PLoS Genet* **2**, e52 (Apr, 2006).

210. E. Birney *et al.*, *Nature* **447**, 799 (Jun 14, 2007).

211. M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, R. A. Young, *Cell* **130**, 77 (Jul 13, 2007).

212. G. A. Calin *et al.*, *Cancer Cell* **12**, 215 (Sep, 2007).

213. S. Efroni *et al.*, *Cell Stem Cell* **2**, 437 (May 8, 2008).

214. L. J. Core, J. J. Waterfall, J. T. Lis, *Science* **322**, 1845 (Dec 19, 2008).

215. A. C. Seila *et al.*, *Science* **322**, 1849 (Dec 19, 2008).

216. T. P. ENCODE, *Nature* **457**, 1028 (Feb 19, 2009).

217. P. Preker *et al.*, *Science* **322**, 1851 (Dec 19, 2008).

218. G. J. Faulkner *et al.*, *Nat Genet* **41**, 563 (May, 2009).

219. J. R. Manak *et al.*, *Nat Genet* **38**, 1151 (Oct, 2006).

220. X. Y. Li *et al.*, *PLoS Biol* **6**, e27 (Feb, 2008).

221. H. He *et al.*, *Genome Res* **17**, 1471 (Oct, 2007).

222. L. W. Hillier *et al.*, *Genome Res* **19**, 657 (Apr, 2009).

223. H. Shin *et al.*, *BMC Biol* **6**, 30 (2008).

224. N. Dutrow *et al.*, *Nat Genet* **40**, 977 (Aug, 2008).

225. B. T. Wilhelm *et al.*, *Nature* **453**, 1239 (Jun 26, 2008).

226. K. Hirota, K. Ohta, *Epigenetics* **4**, 5 (Jan, 2009).

227. K. Hirota *et al.*, *Nature* **456**, 130 (Nov 6, 2008).

228. P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **254**, 1497 (Dec 6, 1991).

229. M. Egholm *et al.*, *Nature* **365**, 566 (Oct 7, 1993).

230. K. Kaihatsu, B. A. Janowski, D. R. Corey, *Chem Biol* **11**, 749 (Jun, 2004).

231. B. A. Janowski *et al.*, *Nat Chem Biol* **1**, 210 (Sep, 2005).

232. F. C. Holstege, U. Fiedler, H. T. Timmers, *EMBO J* **16**, 7468 (Dec 15, 1997).

233. T. Bentin, P. E. Nielsen, *Biochemistry* **35**, 8863 (Jul 9, 1996).

234. X. Zhang, T. Ishihara, D. R. Corey, *Nucleic Acids Res* **28**, 3332 (Sep 1, 2000).

235. R. Beane, S. Gabillet, C. Montaillier, K. Arar, D. R. Corey, *Biochemistry* **47**, 13147 (Dec 16, 2008).

236. R. L. Beane *et al.*, *Biochemistry* **46**, 7572 (Jun 26, 2007).

237. C. Giardina, J. T. Lis, *Science* **261**, 759 (Aug 6, 1993).

238. B. A. Janowski *et al.*, *Nat Chem Biol* **1**, 216 (Sep, 2005).

239. Y. Maida *et al.*, *Nature* **461**, 230 (Sep 10, 2009).

240. B. A. Janowski *et al.*, *Nat Struct Mol Biol* **13**, 787 (Sep, 2006).

241. G. Meister, *Science* **321**, 496 (Jul 25, 2008).

242. S. T. Younger, D. R. Corey, *Chembiochem* **10**, 1135 (May 4, 2009).

243. G. B. Robb, K. M. Brown, J. Khurana, T. M. Rana, *Nat Struct Mol Biol* **12**, 133 (Feb, 2005).

244. S. Y. Berezhna, L. Supekova, F. Supek, P. G. Schultz, A. A. Deniz, *Proc Natl Acad Sci U S A* **103**, 7682 (May 16, 2006).

245. K. Ikeda *et al.*, *J Immunol Methods* **317**, 38 (Dec 20, 2006).

246. S. Till *et al.*, *Nat Struct Mol Biol* **14**, 897 (Oct, 2007).

247. S. Guang *et al.*, *Science* **321**, 537 (Jul 25, 2008).

248. J. C. van Wolfswinkel *et al.*, *Cell* **139**, 135 (Oct 2, 2009).

249. W. Gu *et al.*, *Mol Cell* **36**, 231 (Oct 23, 2009).

250. J. M. Claycomb *et al.*, *Cell* **139**, 123 (Oct 2, 2009).

251. T. Ohrt *et al.*, *Nucleic Acids Res* **36**, 6439 (Nov, 2008).

252. P. T. Nelson, A. G. Hatzigeorgiou, Z. Mourelatos, *RNA* **10**, 387 (Mar, 2004).

253. M. Landthaler *et al.*, *RNA* **14**, 2580 (Dec, 2008).

254. L. Weinmann *et al.*, *Cell* **136**, 496 (Feb 6, 2009).

255. A. H. Ting, K. E. Schuebel, J. G. Herman, S. B. Baylin, *Nat Genet* **37**, 906 (Aug, 2005).

256. D. H. Kim, L. M. Villeneuve, K. V. Morris, J. J. Rossi, *Nat Struct Mol Biol* **13**, 793 (Sep, 2006).

257. D. H. Kim, P. Saetrom, O. Snove, Jr., J. J. Rossi, *Proc Natl Acad Sci U S A* **105**, 16230 (Oct 21, 2008).

258. K. Suzuki *et al.*, *J Biol Chem* **283**, 23353 (Aug 22, 2008).

259. M. Allo *et al.*, *Nat Struct Mol Biol* **16**, 717 (Jul, 2009).

260. P. G. Hawkins, S. Santoso, C. Adams, V. Anest, K. V. Morris, *Nucleic Acids Res* **37**, 2984 (May, 2009).

261. S. Napoli, C. Pastori, M. Magistri, G. M. Carbone, C. V. Catapano, *EMBO J* **28**, 1708 (Jun 17, 2009).

262. C. R. Clapier, B. R. Cairns, *Annu Rev Biochem* **78**, 273 (2009).

263. S. Kundu, C. L. Peterson, *Biochim Biophys Acta* **1790**, 445 (Jun, 2009).

264. K. Raemdonck, R. E. Vandenbroucke, J. Demeester, N. N. Sanders, S. C. De Smedt, *Drug Discov Today* **13**, 917 (Nov, 2008).

265. K. V. Morris, *Epigenetics* **4**, 296 (Jul, 2009).

266. S. Gonzalez, D. G. Pisano, M. Serrano, *Cell Cycle* **7**, 2601 (Aug 15, 2008).

267. H. Kawasaki, K. Taira, K. V. Morris, *Cell Cycle* **4**, 442 (Mar, 2005).

268. B. A. Janowski *et al.*, *Nat Chem Biol* **3**, 166 (2007).

269. L. C. Li *et al.*, *Proc Natl Acad Sci U S A* **103**, 17337 (Nov 14, 2006).

270. P. Huertas, A. Aguilera, *Mol Cell* **12**, 711 (Sep, 2003).

271. K. Yu, F. Chedin, C. L. Hsieh, T. E. Wilson, M. R. Lieber, *Nat Immunol* **4**, 442 (May, 2003).

272. Y. R. Yuan *et al.*, *Mol Cell* **19**, 405 (Aug 5, 2005).

273. K. B. Mullis, F. A. Faloona, *Methods Enzymol* **155**, 335 (1987).

274. M. A. Abbott *et al.*, *J Infect Dis* **158**, 1158 (Dec, 1988).

275. R. K. Saiki *et al.*, *Science* **239**, 487 (Jan 29, 1988).

276. S. L. McKnight, R. Kingsbury, *Science* **217**, 316 (Jul 23, 1982).

277. K. A. Jones, K. R. Yamamoto, R. Tjian, *Cell* **42**, 559 (Sep, 1985).

278. B. S. Zhou, D. R. Beidler, Y. C. Cheng, *Cancer Res* **52**, 4280 (Aug 1, 1992).

279. M. W. Pfaffl, *Nucleic Acids Res* **29**, e45 (May 1, 2001).

280. J. Vandesompele *et al.*, *Genome Biol* **3**, RESEARCH0034 (Jun 18, 2002).

281. P. Kastner *et al.*, *EMBO J* **9**, 1603 (May, 1990).

282. R. J. Taft *et al.*, *Nat Genet* **41**, 572 (May, 2009).

283. B. A. Janowski, J. Hu, D. R. Corey, *Nat Protoc* **1**, 436 (2006).

284. K. V. Morris, S. Santoso, A. M. Turner, C. Pastori, P. G. Hawkins, *PLoS Genet* **4**, e1000258 (Nov, 2008).

285. N. Fong, M. Ohman, D. L. Bentley, *Nat Struct Mol Biol* **16**, 916 (Sep, 2009).

286. J. Laurencikiene, A. M. Kallman, N. Fong, D. L. Bentley, M. Ohman, *EMBO Rep* **7**, 303 (Mar, 2006).

287. D. L. Bentley, *Curr Opin Cell Biol* **17**, 251 (Jun, 2005).

288. P. T. Nelson *et al.*, *RNA* **13**, 1787 (Oct, 2007).

289. J. S. Mattick, *PLoS Genet* **5**, e1000459 (Apr, 2009).

290. S. Mahmoudi *et al.*, *Mol Cell* **33**, 462 (Feb 27, 2009).

291. E. Dolgin, *Scientist* **23**, 34 (May, 2009).

292. J. Han, D. Kim, K. V. Morris, *Proc Natl Acad Sci U S A* **104**, 12422 (Jul 24, 2007).

293. B. P. Lewis, C. B. Burge, D. P. Bartel, *Cell* **120**, 15 (Jan 14, 2005).

294. S. M. Tan-Wong, J. D. French, N. J. Proudfoot, M. A. Brown, *Proc Natl Acad Sci U S A* **105**, 5160 (Apr 1, 2008).

295. J. Gil, D. Bernard, D. Martinez, D. Beach, *Nat Cell Biol* **6**, 67 (Jan, 2004).

296. S. S. Levine *et al.*, *Mol Cell Biol* **22**, 6070 (Sep, 2002).

297. N. Dietrich *et al.*, *EMBO J* **26**, 1637 (Mar 21, 2007).

298. M. J. Moore, N. J. Proudfoot, *Cell* **136**, 688 (Feb 20, 2009).

299. J. M. O'Sullivan *et al.*, *Nat Genet* **36**, 1014 (Sep, 2004).

300. A. Ansari, M. Hampsey, *Genes Dev* **19**, 2969 (Dec 15, 2005).

301. J. C. Dantonel, K. G. Murthy, J. L. Manley, L. Tora, *Nature* **389**, 399 (Sep 25, 1997).

302. W. H. Wu, I. Pinto, B. S. Chen, M. Hampsey, *Genetics* **153**, 643 (Oct, 1999).

303. B. N. Singh, M. Hampsey, *Mol Cell* **27**, 806 (Sep 7, 2007).

304. K. J. Perkins, M. Lusic, I. Mitar, M. Giacca, N. J. Proudfoot, *Mol Cell* **29**, 56 (Jan 18, 2008).

305. V. K. Tiwari *et al.*, *PLoS Biol* **6**, 2911 (Dec 2, 2008).

306. L. L. Wei *et al.*, *Mol Endocrinol* **2**, 62 (Jan, 1988).

307. L. L. Wei *et al.*, *Biochemistry* **26**, 6262 (Sep 22, 1987).

308. L. L. Wei, C. Gonzalez-Aller, W. M. Wood, L. A. Miller, K. B. Horwitz, *Mol Endocrinol* **4**, 1833 (Dec, 1990).

309. J. K. Richer *et al.*, *Breast Cancer Res Treat* **48**, 231 (Apr, 1998).

310. D. M. Cork, T. W. Lennard, A. J. Tyson-Capper, *Breast Cancer Res* **10**, 207 (2008).

311. L. L. Wei, R. Miner, *Cancer Res* **54**, 340 (Jan 15, 1994).

312. L. L. Wei *et al.*, *Mol Endocrinol* **10**, 1379 (Nov, 1996).

313. J. C. Condon, D. B. Hardy, K. Kovaric, C. R. Mendelson, *Mol Endocrinol* **20**, 764 (Apr, 2006).

314. A. Samalecos, B. Gellersen, *Endocrinology* **149**, 5872 (Nov, 2008).

315. J. Lykke-Andersen, M. D. Shu, J. A. Steitz, *Cell* **103**, 1121 (Dec 22, 2000).

316. G. Singh, S. Jakob, M. G. Kleedehn, J. Lykke-Andersen, *Mol Cell* **27**, 780 (Sep 7, 2007).

317. J. Weischenfeldt, J. Lykke-Andersen, B. Porse, *Curr Biol* **15**, R559 (Jul 26, 2005).

318. J. Houseley, D. Tollervey, *Cell* **136**, 763 (Feb 20, 2009).

319. B. P. Brandhorst, E. H. McConkey, *Journal of Molecular Biology* **85**, 451 (1974).

320. B. P. Brandhorst, E. H. McConkey, *Proc Natl Acad Sci U S A* **72**, 3580 (Sep, 1975).

321. S. Pinol-Roma, G. Dreyfuss, *Nature* **355**, 730 (Feb 20, 1992).

322. G. Dreyfuss, M. J. Matunis, S. Pinol-Roma, C. G. Burd, *Annu Rev Biochem* **62**, 289 (1993).

323. A. Aguilera, *Curr Opin Cell Biol* **17**, 242 (Jun, 2005).

324. H. Mehlin, B. Daneholt, U. Skoglund, *Cell* **69**, 605 (May 15, 1992).

325. X. Sun *et al.*, *J Cell Biol* **142**, 1181 (Sep 7, 1998).

326. C. Saguez, J. R. Olesen, T. H. Jensen, *Curr Opin Cell Biol* **17**, 287 (Jun, 2005).

327. P. Sommer, U. Nehrbass, *Curr Opin Cell Biol* **17**, 294 (Jun, 2005).

328. K. M. Reinisch, S. L. Wolin, *Curr Opin Struct Biol* **17**, 209 (Apr, 2007).

329. J. Houseley, J. LaCava, D. Tollervey, *Nat Rev Mol Cell Biol* **7**, 529 (Jul, 2006).

330. J. Houseley, D. Tollervey, *Biochim Biophys Acta* **1779**, 239 (Apr, 2008).

331. D. Bentley, *Curr Opin Cell Biol* **14**, 336 (Jun, 2002).

332. A. Shilatifard, R. C. Conaway, J. W. Conaway, *Annu Rev Biochem* **72**, 693 (2003).

333. G. Bird, D. A. Zorio, D. L. Bentley, *Mol Cell Biol* **24**, 8963 (Oct, 2004).

334. F. Ferre, *PCR Methods Appl* **2**, 1 (Aug, 1992).

335. D. S. Millican, I. M. Bird, *Methods Mol Biol* **105**, 337 (1998).

336. B. A. Kraynack, B. F. Baker, *RNA* **12**, 163 (Jan, 2006).

337. Y. Mizuno *et al.*, *Nucleic Acids Res* **27**, 1345 (Mar 1, 1999).

338. A. Gondor, C. Rougier, R. Ohlsson, *Nat Protoc* **3**, 303 (2008).

339. A. Miele, N. Gheldof, T. M. Tabuchi, J. Dostie, J. Dekker, *Curr Protoc Mol Biol* **Chapter 21**, Unit 21 11 (May, 2006).

340. E. Lieberman-Aiden *et al.*, *Science* **326**, 289 (Oct 9, 2009).

341. J. C. Eissenberg, A. Shilatifard, *Curr Opin Genet Dev* **16**, 184 (Apr, 2006).

342. M. F. Ruh, S. Tian, L. K. Cox, T. S. Ruh, *Endocrine* **11**, 157 (Oct, 1999).

343. X. Liu, F. Jiang, S. Kalidas, D. Smith, Q. Liu, *RNA* **12**, 1514 (Aug, 2006).

344. Q. Liu *et al.*, *Science* **301**, 1921 (Sep 26, 2003).

345. H. W. Wang *et al.*, *Nat Struct Mol Biol* **16**, 1148 (Nov, 2009).

346. T. P. Chendrimada *et al.*, *Nature* **436**, 740 (Aug 4, 2005).

347. Y. Lee *et al.*, *EMBO J* **25**, 522 (Feb 8, 2006).

348. G. B. Robb, T. M. Rana, *Mol Cell* **26**, 523 (May 25, 2007).

349. G. Meister *et al.*, *Curr Biol* **15**, 2149 (Dec 6, 2005).

350. G. Hutvagner, P. D. Zamore, *Science* **297**, 2056 (Sep 20, 2002).

351. J. C. Schwartz *et al.*, *Nat Struct Mol Biol* **15**, 842 (Aug, 2008).

352. S. A. Lelievre, *Biochim Biophys Acta* **1790**, 925 (Sep, 2009).