The Puzzle of RNAs that Target Gene Promoters

Scott T. Younger and David R. Corey*[a]

Few discoveries have had the impact of RNAi. Most researchers who use small RNAs to control gene expression in mammalian cells assume that mRNA will be the target. Recent studies, however, have suggested that small RNAs can also target chromosomal DNA.

Introduction

Small duplex RNAs have become ubiquitous tools for controlling gene expression.[1] When introduced into cells or expressed inside cells, duplex RNAs can target mRNA and reduce gene expression. Gene silencing by small RNAs is robust, potent, and readily achievable by laboratories with modest expertise in working with mammalian cells. Few discoveries have had such a large and immediate impact on the conduct of experimental science.

The common assumption for RNA-mediated gene silencing in mammalian cells is that small RNAs target mRNA. In 2004, however, two papers appeared that challenged this common wisdom.[2, 3] These papers reported that small RNAs could target promoter DNA, induce DNA methylation, and reduce gene expression.

While these findings were intriguing, acceptance of the possibility that RNAs might mediate promoter function has been slow. One problem was that one[3] of the initial two papers was subsequently retracted.[4] On a fundamental scientific level, there was no mechanistic framework for understanding the results. mRNA and chromosomal DNA are very different, how could duplex RNAs recognize both of them? On a practical level, experiments that involve introduction of nucleic acids into cells are notoriously artifact prone.[5] Skeptics could easily dismiss initial reports by assuming that effects were due to interactions between the introduced RNAs and nontarget molecules (also known as off-target effects; these can be caused by unintended interactions with proteins, RNA, or DNA).

In late 2006, the excitement and the skepticism that was caused by promoter-targeted RNAs was further increased by a report that these RNAs could also activate gene expression,[6] a capability that would substantially expand the experimental and therapeutic value of duplex RNA. This claim, however, fit even less well into the established theoretical framework for RNAi.[7]

How Do Promoter-Targeted RNAs Function?

Prior to these initial reports on modulating gene expression by promoter-targeted RNAs, our laboratory had been working on the recognition of chromosomal DNA by synthetic peptide nucleic acid oligomers (PNAs; Scheme 1) that are complementary to transcription start sites.[8] Recognition of chromosomal DNA
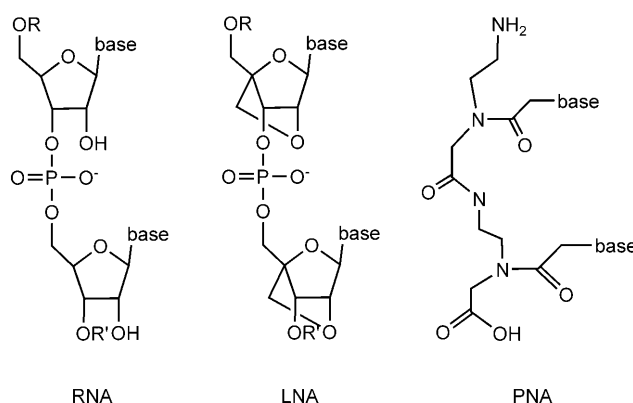
Scheme 1. Chemical structure of RNA compared to structures of synthetic nucleic acid mimics, locked nucleic acid (LNA),[11] and peptide nucleic acid (PNA)[8] that are known to recognize duplex DNA inside cells. Antigene PNAs and LNAs are introduced into cells as single strands, and antigene RNAs are duplexes. PNAs and LNAs provide useful benchmarks for evaluating the activities of agRNAs and considering potential mechanisms of action.

by PNAs would necessarily follow much different rules than recognition of DNA by duplex RNA because PNAs are single stranded and have an unnatural amide backbone. Nevertheless, it was straightforward for us to use our established experimental systems to investigate the potential for duplex RNA to target chromosomal DNA.

We designed duplex RNAs to target the promoter of the progesterone receptor (PR) gene.[9] We refer to these RNAs as antigene RNAs (agRNAs) to distinguish them from duplex RNAs that are designed to target mRNA and to conform to traditional chemical nomenclature for synthetic oligomers that are intended to interact with chromosomal DNA.

We chose to target PR because we had already shown that its expression could be reduced by single-stranded PNAs,[8] and its promoter region was well characterized.[10] This last point is critical for experiments with RNAs that target DNA. Experimenters must be confident that the transcription start site

[a] S. T. Younger, Prof. D. R. Corey
Department of Pharmacology, UT Southwestern Medical Center
Dallas, TX 75390-9041 (USA)
Fax: (+ 1) 214-645-6067
E-mail: david.corey@utsouthwestern.edu

has been accurately determined and that observed effects are not due to interactions with mRNA. Recently, we have also shown that single-stranded locked nucleic acids (LNAs) can recognize chromosomal DNA inside cells.[11] This observation reinforces the suggestion that the PR promoter is susceptible to binding by synthetic oligomers and is a productive model system for examining agents designed to modulate gene expression.

Our experiments revealed a dramatic reduction of PR mRNA levels and protein expression upon addition of agRNAs.[9] In contrast to experiments with PNAs or LNAs, that required extensive optimization, the first experiments with agRNAs revealed potent inhibition, and subsequent experiments have demonstrated that the results are easy to reproduce. This robustness hints at the possibility that the activity of our agRNAs reflects natural mechanisms for promoter-targeted RNAs. Extensive control experiments suggested that our agRNAs were interacting with their intended target sequences and support the conclusion that we were investigating a novel phenomenon.

We screened through a large number of agRNAs and identified several that are capable of inhibiting PR gene expression.[9] In the course of these experiments, we also identified RNAs that reproducibly caused a 25–50% increase in PR gene expression. While such relatively small changes were far from definitive, they did suggest that some small RNAs might be able to act like protein transcription factors and enhance gene expression. Up to this point our experiments had been carried out in a cell line with high expression of our target gene. We reasoned that using a cell line that expressed low levels of PR would allow increases in gene expression to be observed more easily.

After testing agRNAs in a cell line with low background PR expression it became clear that they had the potential to increase gene expression many-fold above background.[12] We identified several agRNAs capable of activating PR expression, whereas multiple mismatch and scrambled control duplex RNAs did not affect gene expression. These data suggested that agRNAs could activate gene expression.

Unlike traditional siRNAs, agRNAs are not complementary to mRNA. What molecules might agRNAs be associating with to produce such varied effects on gene expression? One hypothesis is that agRNAs bind directly to chromosomal DNA (Figure 1A). While conceptually simple, there is no known cellular machinery for promoting this recognition. An alternative explanation is that the agRNAs bind to nascent RNA transcripts that overlap gene promoters (Figure 1B).

These hypotheses were first described by investigations using yeast as a model system.[13] RNAi proteins had been implicated in heterochromatin formation in fission yeast through the RNA initiation of transcriptional gene silencing (RITS) complex formation.[14] The yeast RITS complex is comprised of Chp1, Tas3, and Ago1. Of these proteins, only Ago1 has clear mammalian homologues. Mammalian cells contain four Ago genes (Ago1–Ago4). Of these genes, Ago2 has been described as the catalytic engine for RNAi, and the functions of Ago1, Ago3, and Ago4 remain unclear.[15, 16]
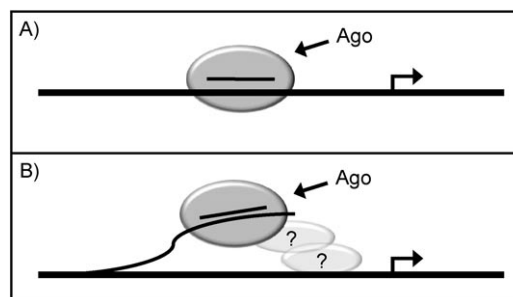
Figure 1. Possible mechanism for agRNA. A) agRNAs could directly recognize chromosomal DNA. B) agRNAs could recognize ncRNA transcripts originating within the target gene promoter. Argonaute family proteins (Ago) are known to mediate agRNA function. Other, unidentified proteins could bridge a ncRNA and promoter DNA to affect transcription and gene expression.

We investigated the potential involvement of Ago proteins in gene regulation by agRNAs in mammalian cells. We observed that Ago2, and to a lesser extent Ago1 appeared to be necessary for silencing by agRNAs.[17] Ago1 involvement was reported for gene silencing by Rossi and co-workers,[18] whereas Li and co-workers implicated Ago2 in RNA-mediated gene activation.[6] The Ago proteins are known to promote RNA-mediated recognition of RNA; this further suggests a model in which agRNAs bind nascent RNA transcripts rather than chromosomal DNA.

Whereas there were no reports of RNA being produced from the PR gene promoter, recent studies had revealed that most of the genome is transcribed.[19–22] Complex networks of noncoding RNAs (ncRNAs) overlap gene promoters as well as other regions of the genome. These ncRNAs are transcribed in both the sense and antisense direction relative to the genes in which they overlap. These findings led us to search for potential ncRNAs that overlapped the PR gene promoter.

By using RT-PCR we detected RNA species that originate from the PR gene promoter and 5′-RACE identified a ncRNA that is transcribed through the promoter in the antisense direction relative to the PR gene.[23] We were unable to detect any RNA species being transcribed upstream of the previously determined[10] transcription start site for PR mRNA in the sense direction relative to the gene. The antisense transcript was, therefore, the only known candidate for recognition by our agRNAs. Having identified a candidate RNA target for our agRNAs, we used a battery of techniques to investigate the possible mechanistic role of this ncRNA in our observations.

Our experiments revealed several lines of evidence implicating the antisense transcript in the action of agRNAs: 1) We developed a biotin–avidin purification assay in which the agRNA is biotinylated on either strand of the duplex and transfected into cells. The bound nucleic acid is isolated by using avidin-coated beads and analyzed by using RT-PCR. We observed association of the agRNA with the antisense ncRNA transcript and no association with chromosomal DNA; 2) We found that agRNAs could induce association of Ago proteins with the ncRNA transcript by using RNA immunoprecipitation; 3) re-

ducing the levels of the antisense ncRNA reversed the activity of activating agRNAs.

Taken together, these data suggest a model in which the agRNA recognizes the antisense ncRNA in close proximity to the chromosome and transcriptionally modulates gene expression (Figure 2). Supporting this hypothesis, chromatin immunoprecipitation for RNA Pol II reveals increased and decreased transcription for activating and silencing agRNAs, respectively.
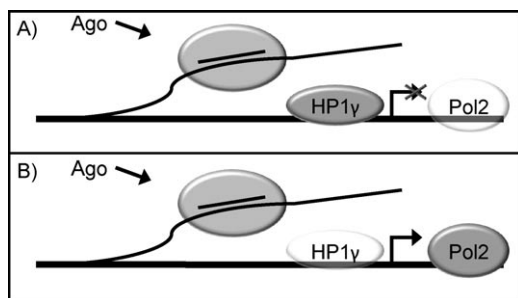

Figure 2. Interactions of duplex RNAs at the PR promoter during silencing or gene activation. A) Silencing agRNAs recruit both AGO and HP1γ to the target gene promoter in cells that normally express a high level of PR. This results in decreased levels of RNA polymerase II on the target gene. B) Activating agRNAs recruit AGO to the target gene promoter and increase RNA polymerase II levels on the target gene in cells that normally express a low level of PR.

Our studies,[23] and those of others[24, 25] have suggested that ncRNAs transcribed through gene promoters can be utilized to regulate gene transcription by using synthetic small duplex RNA, but why are noncoding RNAs transcribed at gene promoters? What is the natural function of these ncRNAs?

The Noncoding RNA world

To begin to answer these questions, it is necessary to understand that transcription is much more complicated than has been appreciated previously. Traditionally, transcription of a gene is pictured as synthesis of an mRNA (Figure 3 A). Slightly
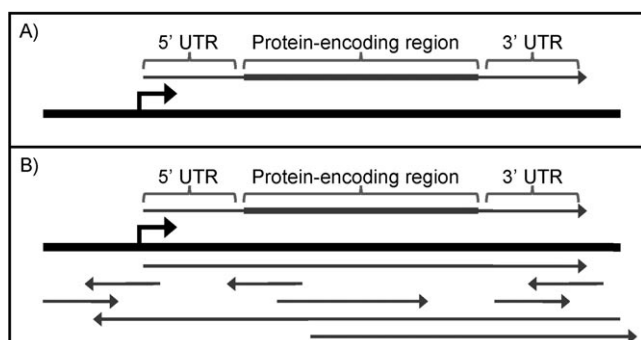
Figure 3. Models of RNA expression. A) Traditional model: an mRNA is synthesized and translated into protein. B) Emerging model: Many overlapping RNAs are synthesized alongside the protein-encoding mRNA. The overall effect of these RNAs on transcription and translation of mRNA are only beginning to be understood.

over 1% of the human genome is transcribed into mRNA and this RNA has been the focus of most experimental studies. Recent studies have made it clear, however, that greater than 80% of the genome is transcribed into RNA.[19–22] Of special significance for the mechanism of agRNAs, a substantial fraction of all genes have ncRNA transcripts that overlap their promoters.[26, 27]

Instead of picturing a simple mRNA being transcribed, it is more accurate to envision the mRNA being just one of many overlapping RNAs (Figure 3 B). Our understanding of the function of these RNAs, or whether most have any function at all, is in its infancy. Several recent reports, however, have investigated various functional aspects of ncRNAs and their involvement in regulating gene transcription (Figure 4).
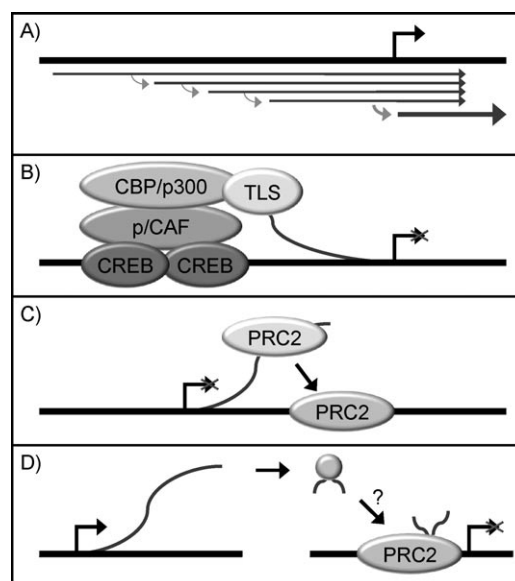
Figure 4. Different mechanisms of gene regulation by ncRNAs. A) Transcription of an upstream ncRNA alters chromatin and allows transcription of subsequent ncRNAs. The end result is increased transcription of the associated gene. B) Protein binding to a ncRNA that is produced from the gene promoter causes allosteric modifications that result in local transcriptional silencing. C) Transcriptional-silencing complexes bind to the ncRNA in close proximity to the chromosome and are transferred to the DNA, resulting in transcriptional silencing. D) Transcriptional-silencing complexes bind to the ncRNA and are recruited to distant genomic locations by unclear mechanisms.

One group has found that transcription through a gene's promoter modifies the chromatin in a manner that prepares the gene for transcriptional activation.[28] By using the fission yeast *Schizosaccharomyces pombe* as a model system, Hirota et al. were able to demonstrate that a cascade of four ncRNAs transcribed through the promoter of the *fbp1*⁺ gene prepare the locus for activation (Figure 4 A). The authors report that transcription of the most upstream ncRNA prepares the chromatin for binding of factors that induce transcription of the following ncRNA within the cascade. This "stepwise" process results in robust activation of *fbp1*⁺ during glucose starvation. Removal of the ncRNAs via insertion of a premature transcription terminator sequence within the locus was sufficient to

prevent chromatin remodeling and transcriptional activation. This report highlights at least one functional importance of transcription within gene promoters.

Additional reports have focused on the functional roles of the actual ncRNA molecules. Wang et al. showed in human cells that the *cyclin D1* gene promoter produces a ncRNA in response to DNA damage signals.[29] The nascent ncRNA recruits protein factors that repress the *cyclin D1* gene (Figure 4B). A consensus sequence within the ncRNA serves as a ligand for the RNA-binding protein TLS (for translocated in liposarcoma). Upon binding, it is suggested that an allosteric modification in TLS results in an inhibitory effect on the histone acetyltransferases CBP (CREB-binding protein) and p300, thus silencing transcription of the *cyclin D1* gene. The authors propose a general model in which induced ncRNAs can serve as ligands for co-regulatory proteins. The binding of the ncRNA to the co-regulator would result in activation (or inactivation) of the protein; this would cause changes in transcription.

Interestingly, regulation of transcription by ncRNAs does not seem to be restricted to gene promoters. The idea of ncRNAs acting as ligands for co-regulators has been extended to X-chromosome inactivation (XCI). It has long been known that expression of a ncRNA from the X chromosome, termed *Xist*, is responsible for the initiation of XCI.[30,31] Zhao et al. have demonstrated that a short ncRNA expressed from within *Xist* functions to recruit the Polycomb complex, PRC2, to the X chromosome (Figure 4C).[32] They report that depletion of either PRC2 or the ncRNA results in failure to initiate XCI.

The investigations described above have all elucidated the function of ncRNAs acting in *cis* with respect to their own genomic origin. However, there are some examples of ncRNAs that silence transcription by acting in *trans* (Figure 4D). HOTAIR is a long ncRNA that is transcribed from the human *HOXC* locus (on chromosome 12) that represses transcription of the *HOXD* locus (on chromosome 2).[33] Rinn et al. have reported that depletion of the HOTAIR transcript by using multiple siRNAs results in significant increases in expression of the *HOXD* cluster but has no effect on the *HOXC* cluster itself. The authors also suggest involvement of the same Polycomb complex (PRC2) that was reported to be involved in ncRNA-mediated X-chromosome inactivation.

Another example of an ncRNA acting in *trans* to silence transcription is the *Air* ncRNA. The *Air* transcript is expressed from paternally inherited chromosomes and is involved in imprinting (or silencing) of the proximal genes on the chromosome.[34] The *Air* transcript does not affect expression of the same locus on the maternally inherited chromosome. A recent report from Nagano et al. showed that the ncRNA physically interacts with the promoter of one of the imprinted genes, *Slc22a3*.[35] *Air* is considered to work in *trans* because the gene and the ncRNA are positioned more than 100 kb apart and are transcribed in divergent orientation. The report suggests that the *Air* ncRNA recruits the histone methyltransferase G9a to the *Slc22a3* promoter. However, evidence also suggests that the Air ncRNA uses different mechanisms to imprint other genes in the locus.

Conclusions

It is becoming apparent that ncRNAs play important roles in diverse cellular processes, and the pool of interesting ncRNAs is increasing at a rapid pace. However, because ncRNAs inherently have no coding potential, it remains difficult to accurately predict their function. Much more work will be required before we can understand how these RNAs contribute to regulating cellular processes.

Many open mechanistic questions remain unanswered and will likely be a rich area of research for many years. For ncRNAs in general, little is known about their function or how they interact with chromosomal DNA, proteins, or mRNA. For agRNAs, the precise molecular details for how they can activate expression in one context and silence expression in another remain to be elucidated. Perhaps the biggest question is whether endogenous small RNAs, for example microRNAs, possess the ability to recognize gene promoters.

For chemists, the ability to either activate or inhibit gene expression by using synthetic agRNAs that are complementary to gene promoters hints that noncoding RNAs provide a large and virtually untapped reservoir of molecular targets. It is likely that we have glimpsed only a small fraction of the noncoding RNA world. Chemical approaches will contribute to defining the structure of these RNAs, their biological functions, and the potential for using them to manipulate the expression of genes.

[1] J. K. Watts, G. F. Deleavey, M. J. Damha, *Drug Discovery Today* **2008**, *13*, 842–855.

[2] K. V. Morris, S. W. Chan, S. E. Jacobsen, D. J. Looney, *Science* **2004**, *305*, 1289–1292.

[3] H. Kawasaki, K. Taira, *Nature* **2004**, *431*, 211–217.

[4] K. Taira, *Nature* **2006**, *441*, 1176.

[5] A. L. Jackson, J. Burchard, J. Schelter, B. N. Chau, M. Cleary, L. Lim, P. S. Linsley, *RNA* **2006**, *12*, 1179–1187.

[6] L. Li, S. T. Okino, H. Zhao, D. Pookot, S. Urakami, H. Enokida, R. Dahiya, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17337–17342.

[7] E. Check, *Nature* **2007**, *448*, 855–858.

[8] B. A. Janowski, K. Kaihatsu, K. E. Huffman, J. C. Schwartz, R. Ram, D. B. Hardy, C. R. Mendelson, D. R. Corey, *Nat. Chem. Biol.* **2005**, *1*, 210–215.

[9] B. A. Janowski, K. E. Huffman, J. C. Schwartz, R. Ram, D. B. Hardy, D. S. Shames, J. D. Minna, D. R. Corey, *Nat. Chem. Biol.* **2005**, *1*, 216–222.

[10] P. Kastner, A. Krust, B. Turcotte, U. Stropp, L. Tora, H. Gronemeyer, P. Chambon, *EMBO J.* **1990**, *9*, 1603–1614.

[11] R. Beane, S. Gabillet, C. Montallier, K. Arar, D. R. Corey, *Biochemistry* **2008**, *47*, 13147–13149.

[12] B. A. Janowski, S. T. Younger, D. B. Hardy, R. Ram, K. E. Huffman, D. R. Corey, *Nat. Chem. Biol.* **2007**, *3*, 166–173.

[13] M. Buhler, A. Verdel, D. Moazed, *Cell* **2006**, *125*, 873–886.

[14] A. Verdel, S. Jia, S. Gerber, T. Sugiyama, S. Bygi, S. I. Grewal, D. Moazed, *Science* **2004**, *303*, 672–676.

[15] G. Meister, M. Landthaler, A. Patkaniowska, Y. Dorsett, G. Teng, T. Tuschl, *Mol. Cell* **2004**, *15*, 185–197.

[16] J. Liu, M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J. J. Song, S. M. Hammond, L. Joshua-Tor, G. J. Hannon, *Science* **2004**, *305*, 1437–1441.

[17] B. A. Janowski, K. E. Huffman, J. C. Schwartz, R. Ram, R. Nordsell, D. S. Shames, J. D. Minna, D. R. Corey, *Nat. Struct. Biol.* **2006**, *13*, 787–792.

[18] D. H. Kim, L. M. Villeneuve, K. V. Morris, J. J. Rossi, *Nat. Struct. Biol.* **2006**, *13*, 793–797.

[19] RIKEN Group and FANTOM Consortium, *Science* **2005**, *309*, 1564–1566.

[20] The Encode Project Consortium, *Nature* **2007**, *447*, 799–816.

[21] T. R. Gingeras, *Genome Res.* **2007**, *17*, 682–690.

[22] P. P. Amaral, J. S. Mattick, *Mamm. Genome* **2008**, *19*, 454–492.

[23] J. C. Schwartz, S. T. Younger, N. B. Nguyen, D. B. Hardy, D. R. Corey, *Nat. Struct. Mol. Biol.* **2008**, *15*, 842–848.

[24] J. Han, D. Kim, K. V. Morris, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12422–14227.

[25] K. V. Morris, S. Santoso, A. M. Turner, C. Pastori, P. G. Hawkins, *PLoS Genet.* **2008**, ePub.

[26] R. Yelin, D. Dahary, R. Sorek, E. Y. Levanon, O. Goldstein, A Shoshan, A. Diber, S. Biton, Y. Tamir, R. Khosravi, S. Nemzer, E. Pinner, S. Walach, J. Bernstein, K. Savitsky, G. Rotman, *Nat. Biotechnol.* **2003**, *21*, 379–385.

[27] Y. He, B. Vogelstein, V. E. Velculescu, N. Papadopoulos, K. W. Kinzler, *Science* **2008**, *322*, 1855–1857.

[28] K. Hirota, T. Mioshi, K. Kogou, C. S. Hoffman, T. Shibata, K. Ohta, *Nature* **2008**, *456*, 130–134.

[29] X. Wang, S. Arai, X. Song, D. Reichart, K. Du, G. P. Pascual, G. Tempst. M. G. Rosenfeld, C. K. Glass, R. Kurokawa, *Nature* **2008**, *454*, 126–130.

[30] C. J. Brown, B. D. Hendrich, J. L. Rupert, R. G. Lafreniere, Y. Xing, J. Lawrence, H. F. Willard, *Cell* **1992**, *71*, 527–542.

[31] G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, *Nature* **1996**, *379*, 131–137.

[32] J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, *Science* **2008**, *322*, 750–756.

[33] J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, H. Y. Chang, *Cell* **2007**, *129*, 1311–1323.

[34] A. Wutz, D. P. Barlow, *Nature* **1997**, *389*, 745–749.

[35] T. Nagano, J. A. Mitchell, L. A. Sanz, F. M. Pauler, A. C. Ferguson-Smith, R. Feil, P. Fraser, *Science* **2008**, *322*, 1717–1720.