

Long Noncoding RNA, Polycomb, and the Ghosts Haunting *INK4b-ARF-INK4a* Expression

Francesca Aguiló^{1,2}, Ming-Ming Zhou¹, and Martin J. Walsh^{1,2,3}

Abstract

Polycomb group proteins (PcG) function as transcriptional repressors of gene expression. The important role of PcG in mediating repression of the *INK4b-ARF-INK4a* locus, by directly binding to the long noncoding RNA (lncRNA) transcript antisense noncoding RNA in the *INK4* locus (*ANRIL*), was recently shown. *INK4b-ARF-INK4a* encodes 3 tumor-suppressor proteins, p15^{INK4b}, p14^{ARF}, and p16^{INK4a}, and its transcription is a key requirement for replicative or oncogene-induced senescence and constitutes an important barrier for tumor growth. *ANRIL* gene is transcribed in the antisense orientation of the *INK4b-ARF-INK4a* gene cluster, and different single-nucleotide polymorphisms are associated with increased susceptibility to several diseases. Although lncRNA-mediated regulation of *INK4b-ARF-INK4a* gene is not restricted to *ANRIL*, both polycomb repressive complex-1 (PRC1) and -2 (PRC2) interact with *ANRIL* to form heterochromatin surrounding the *INK4b-ARF-INK4a* locus, leading to its repression. This mechanism would provide an increased advantage for bypassing senescence, sustaining the requirements for the proliferation of stem and/or progenitor cell populations or inappropriately leading to oncogenesis through the aberrant saturation of the *INK4b-ARF-INK4a* locus by PcG complexes. In this review, we summarize recent findings on the underlying epigenetic mechanisms that link PcG function with *ANRIL*, which impose gene silencing to control cellular homeostasis as well as cancer development. *Cancer Res*; 71(16); 5365–9. ©2011 AACR.

Introduction

Polycomb group (PcG) genes were first identified in *Drosophila melanogaster* as regulators of anterior-posterior body patterning through modulation of *Hox* gene expression and were subsequently described as a large group of proteins involved in transcriptional repression (1). In mammals, several highly conserved PcG proteins form 2 large macromolecular complexes classified as polycomb repressive complex-1 (PRC1) and -2 (PRC2), which induce gene silencing via histone modification, and which ultimately affect the chromatin structure. The PRC1 complex includes BMI1, mPh1/2, Pc/Chromobox (CBX), and the ubiquitin E3 ligase RING1A/B, which monoubiquitylates histone 2A lysine 119 (H2AK119ub1) and participates in the maintenance of silent chromatin (2). Mammalian PRC1 may exhibit additional contextual specificity by selecting one of the multiple Pc/CBX homologs. These chromobox proteins confer distinct subchromosomal distribution and dynamic patterns to the complex and have varying

affinities for their methylated substrates (3). The core of the PRC2 complex is composed of EED, SUZ12, and the histone lysine methyltransferases EZH1/2, which catalyze di- and trimethylation of lysine 27 of histone H3 (H3K27me2 and H3K27me3), initiating repression of target genes (4). In the presence of EZH1, the PRC2 complex is also responsible for general monomethylation at H3K27 (H3K27me1) across the genome. This signature may reflect the more constitutive monomethylation state for H3K27 and may prime PRC2-containing EZH2 for more robust silencing. H3K27me3 provides a recruitment site for PRC1 through binding by the chromodomain of the Pc protein (5), which in turn permits H2AK119ub1 by RING1A/B (2).

Changes in chromatin structure can cause aberrant gene expression patterns and genomic instability, giving rise to a transformed cell phenotype and malignant outgrowth. Therefore, proteins that control chromatin organization constitute key players in cancer molecular pathogenesis. EZH2, a PRC2 member, has been shown to be highly expressed in prostate and breast cancer and amplified in a variety of other cancers, linked with the general loss of key tumor-suppressor genes and poor patient survival (6). As for PRC1, overexpression and amplification of BMI1 has been shown in multiple tumor types, including medulloblastoma and mantle cell lymphoma. In addition, high levels of CBX7 are found in germinal center-derived follicular lymphoma and correlated with clinical stage and lymph node metastasis in gastric tumors. Conversely, loss of CBX7 has been linked with prevalence of other tumor types such as thyroid cancer, suggesting an alternative role for CBX7 to direct PRC1 function under diverse cellular

Authors' Affiliations: Departments of ¹Structural and Chemical Biology, ²Pediatrics, and ³Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York

Corresponding Author: Martin J. Walsh, Departments of Structural and Chemical Biology, Pediatrics, and Genetic and Genomic Sciences, Mount Sinai School of Medicine, One Gustave L. Levy Pl., New York, NY 10029. Phone: 212-241-9714; Fax: 212-426-1972; E-mail: martin.walsh@mssm.edu

doi: 10.1158/0008-5472.CAN-10-4379

©2011 American Association for Cancer Research.

contexts. Interestingly, both BMI and CBX7 have been directly linked to the repression of the *INK4b-ARF-INK4a* tumor-suppressor locus, maintaining the balance between cell proliferation and senescence (7, 8).

Although it is clear that the assembly of PRC1/2 at specific chromosomal locations results in transcription silencing, the mode by which PRC1/2s are targeted to specific genetic loci remains poorly defined. In *D. melanogaster*, PcG targeting to chromatin is accomplished to a greater degree by specific DNA-binding proteins, such as *Zeste*, *Pipsqueak*, and *Pho*, which bind to *cis*-regulatory elements termed polycomb-response elements (PRE) (9). Is there a comparable DNA-binding regulatory mechanism in mammals, or is the absence of a concise DNA-binding model a manifestation of multiple mechanisms that may fine-tune PRC activity? The presence of PREs in mammals remains controversial, and new evidence shows that certain lncRNAs are physically linked with PcG complexes to target specific genomic regions. In contrast to the ubiquitously expressed chromatin-modifying proteins, lncRNAs are differentially expressed; this indicates that lncRNAs may play a major role in programming the epigenome by cooperating and coordinating with PcG complexes to impose chromatin states in a dynamic manner (10).

lncRNAs are transcripts longer than 200 nucleotides and lack protein-coding capacity. Although ncRNAs have been referred to ironically as "dark matter" or "nonproductive," they modulate a wide repertoire of biological functions (11). For instance, lncRNAs X inactive-specific transcript (*Xist*), *Tsix*, *Xite*, and *RepA*, are responsible for the 3 steps of random X chromosome inactivation, including counting of the X-to-autosome ratio, choice of which X chromosome to inactivate, and silencing of the inactive X (12). In addition to regulating physiologic functions, ncRNAs are also involved in disease development. Recently, it has been described that overexpression of *HOTAIR* is able to accumulate the PRC2 complex to execute gene silencing, alternate genome-wide H3K27me3 patterns, reset global gene expression profiles, activate metastasis-associated genes, and promote breast cancer metastasis (13). Similarly, the ncRNAs *MALAT1*, *DLK1-GTL2*, and *H19* are involved in the development of lung adenocarcinomas, neuroblastoma, Wilms tumors, and breast cancer, among others (14, 15).

***ANRIL* and the *INK4b-ARF-INK4a* Locus**

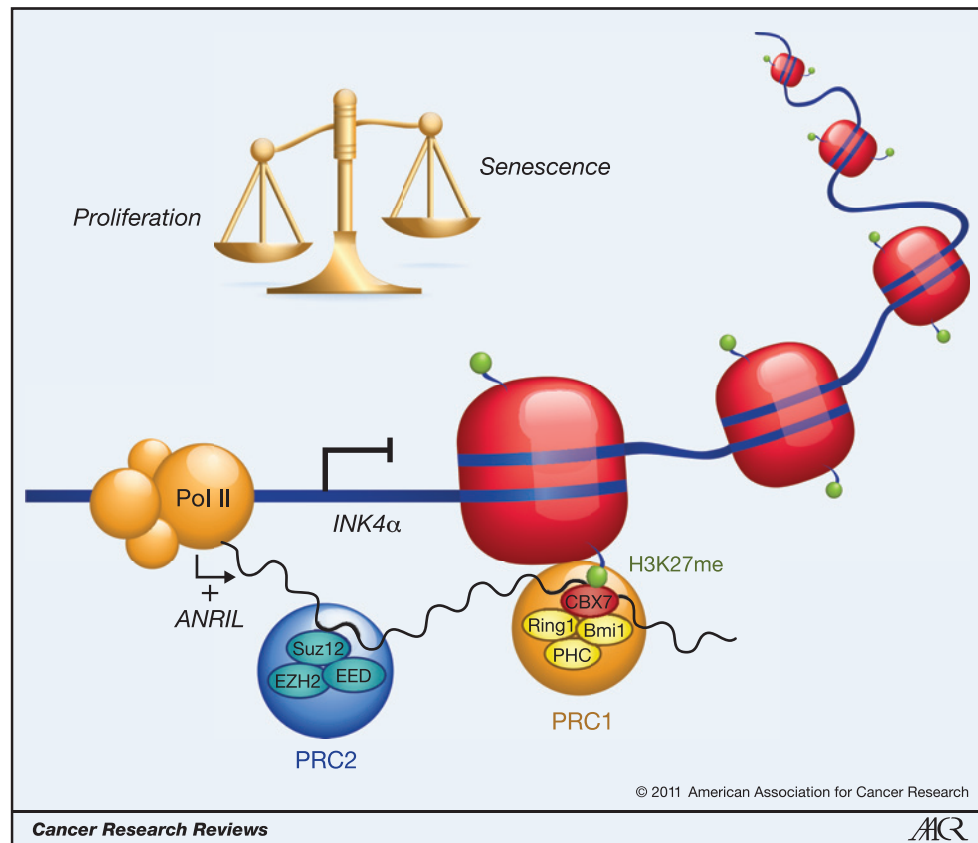
The *INK4b-ARF-INK4a* locus located on the human chromosome 9p21 encodes 3 critical tumor suppressors, p15^{INK4b}, p14^{ARF} (p19^{ARF} in mice), and p16^{INK4a}, which play a central role in cell-cycle inhibition, senescence, and stress-induced apoptosis (16). Although p15^{INK4b} and p16^{INK4a} are often codeleted in tumors, p15 is believed to act as a backup for p16. The *INK4* proteins, p15^{INK4b} and p16^{INK4a}, target cyclin-dependent kinase (CDK) 4/6, prohibiting the binding of these kinases to D-type cyclins and, as a consequence, inhibiting CDK4/6-mediated phosphorylation of retinoblastoma family members. In contrast, p14^{ARF} binds to MDM2 and promotes its degradation, resulting in p53 activation and cell-cycle

arrest. Members of the *INK4b-ARF-INK4a* cluster are key effectors of oncogene-induced senescence and are induced during aging and in premalignant lesions, limiting tumor progression. Therefore, expression of the *INK4b-ARF-INK4a* locus is tightly controlled, and PcG complexes are required to initiate and maintain the silenced state (16, 17). Actually, in multiple cell types, the PRC1 proteins BMI1, PCGF1, PCGF2/MEL18, CBX2, CBX7, CBX8, and RING1B, and the PRC2 proteins EED, SUZ12, and EZH2 have been shown to directly bind to and repress the *INK4b-ARF-INK4a* locus (8, 18, 19). Mechanisms that try to unify our understanding of how PRCs locate their chromosomal targets have begun to elucidate the role of lncRNAs, mainly *ANRIL* but also *Mov10*, to tether both PRC1 and PRC2 to guide *INK4b-ARF-INK4a* locus silencing (20–22). For instance, a recent investigation by Kotake and colleagues revealed that *ANRIL* is targeted by PRC2 to the *INK4b-ARF-INK4a* locus (23). Specifically, depletion of *ANRIL* disrupts the SUZ12 binding to the p15^{INK4b} locus, increases the expression of p15^{INK4b}, but not p14^{ARF} or p16^{INK4a}, and inhibits cellular proliferation. On the other hand, our previous study showed that the PRC1 component CBX7 is tethered by *ANRIL* to the H3K27 methylation mark of the *INK4b-ARF-INK4a* gene cluster to control cellular lifespan (Fig. 1; ref. 24). This topic is discussed further in this review.

The *ANRIL* gene spans a region of 126.3 kb and is transcribed as a 3,834-bp mRNA in the antisense orientation of the *INK4b-ARF-INK4a* gene cluster. The first intron of the *ANRIL* gene overlaps with the 2 exons of p15^{INK4b} and maintains the silencing state of this gene (20). On the basis of EST assembly, the *ANRIL* gene contains 19 exons, many of them consisting entirely of LINE, SINE, and *Alu* repetitive elements (25). The 5' end of the first exon of the *ANRIL* gene is located –300 bp upstream of the transcription start site (TSS) of the p14^{ARF} gene, suggesting that these 2 genes may share a bidirectional promoter. Hence, the expression of *ANRIL* with p14^{ARF} is coordinated both in physiologic and in pathologic conditions. For instance, binding of CCCTC-binding factor (CTCF) in the CpG island overlapping the *ANRIL*-p14^{ARF} promoters protects against DNA methylation, a requirement to maintain the *INK4b-ARF-INK4a* locus in a poised conformation (26). Alternative spliced *ANRIL* variants have been reported, being the transcripts that contain the external exons (from exon 1 to 3 or from exon 13 to 19), more abundant than those containing the internal exons (from exon 4 to 12), which are circular (27).

Disease genome-wide association studies revealed that *ANRIL* is located in a genetic susceptibility locus associated with several diseases, including coronary artery disease (CAD), intracranial aneurysm, type 2 diabetes, and several cancers, such as glioma, basal cell carcinoma, nasopharyngeal carcinoma, and breast cancer (reviewed by ref. 28). Many single-nucleotide polymorphisms (SNP) in this locus alter *ANRIL* structure (27) and *ANRIL* gene expression (29, 30), mediating the susceptibility to a variety of chronic diseases and cancer predisposition. For example, genotypes harboring the risk allele SNP rs10757278 are associated with a reduced expression of p16^{INK4a}, p15^{INK4b}, p14^{ARF}, and *ANRIL* in peripheral blood T cells, which corresponds to an increased risk of

Figure 1. Model of *ANRIL* lncRNA-mediated p16^{INK4a} gene repression by the PRC complex. The nascent *ANRIL* lncRNA is transcribed by RNA polymerase II at the TSS in the p16^{INK4a} gene and associates with Suz12 to recruit PRC2 complex and initiate H3K27me3. Subsequently, PRC1 is recruited by *ANRIL* via direct binding with CBX7 chromodomain, providing another docking site to bind H3K27me3 and allowing H2AK119ub1, which in turn results in the maintenance of epigenetic repression. As a consequence of the *INK4b-ARF-INK4a* locus silencing, cell proliferation is triggered, whereas senescence is inhibited.



CAD (29). More recently, Cunningham and colleagues (30) found that, besides the SNP rs10757278, the SNP rs1333045 is also involved in the susceptibility to this disease, whereas no correlation of these SNPs with p16^{INK4a} or p15^{INK4b} expression was found. Finally, a relationship between the structure of *ANRIL* and disease susceptibility has been reported, correlating both circular and linear *ANRIL* isoforms proximal to the *INK4b-ARF-INK4a* locus with CAD susceptibility (27).

Transcriptional Silencing of p16^{INK4a} by CBX7 and *ANRIL*

Our recent study provides a molecular mechanism of how lncRNA transcripts functionally coordinate with chromatin-associated factors that modify and interact with H3K27 methylation (24). The expression of *ANRIL*, CBX7, and EZH2 is coordinated and elevated in preneoplastic and neoplastic prostate epithelium tissues, which coincides with a decrease in p16^{INK4a} expression. *ANRIL* expression positively correlates with higher binding affinity of CBX7, EZH2, and RNA polymerase II to the p16^{INK4a} and p14^{ARF} promoters. However, there is a negative correlation of RNA polymerase II occupation within the region of the locus separating the *ANRIL* and p16^{INK4a} TSS, suggesting a highly dynamic function of RNA polymerase II on the opposing DNA strand, which may negatively impact p16^{INK4a} transcription. Overexpression of transcripts antisense to the *ANRIL* lncRNA causes an increase

in the expression of p16^{INK4a}, which coincides with a reduction in the CBX7 and EZH2 binding at the p16^{INK4a} TSS, indicating that *ANRIL* and *ANRIL*-associated factors have a critical role in the repression of the *INK4b-ARF-INK4a* locus. Moreover, H3K27me3 is reduced when *ANRIL* is knocked down, whereas H3K4m3 exhibits little change.

To further investigate the relationship between *ANRIL* and PRC1/2, an RNA-chromatin immunoprecipitation (ChIP) assay was done, and it confirmed that CBX7 stably associates with *ANRIL* *in vivo*. Consistently, after treatment of cell nuclei with ribonuclease A, the binding of CBX7 and other PcG components, such as PHC2, Bmi1, and Suz12, is dramatically reduced, as well as the levels of the histone marks associated with PcG-mediated repression, H3K27me3 and H2AK119ub1. Treatment of cell nuclei with the transcriptional inhibitor α amanitin depletes CBX7 along the *INK4b-ARF-INK4a* locus, indicating that the RNA associated with CBX7, *ANRIL*, is a nascent transcript generated by RNA polymerase II.

To gain insight into the interaction between CBX7 and *ANRIL*, an RNA electrophoretic mobility shift assay was done, and it showed that CBX7 binds with *ANRIL* through its chromodomain. Interestingly, CBX7 also uses this domain to bind to H3K27me. To elucidate the molecular interplay between the binding of the CBX7 chromodomain to H3K27me and to RNA, the 3-dimensional structure for these interactions was solved and revealed that the CBX7 chromodomain employs distinct regions and residues for binding H3K27me

(F11, W32, and W35) or RNA (E14, R17, K31, and K33). A point mutation on their individual binding sites further supported the results from the structure-guided analysis.

In an attempt to understand how the CBX7 interaction with H3K27me and RNA has an impact on replicative lifespan, different mutants were assayed in mouse embryonic fibroblasts or in the lung cell line IMR90. During serial passage of these cell lines, p16^{INK4a}-mediated senescence is inversely coupled with *ANRIL* expression, an effect dependent on CBX7 H3K27-binding activity and, with lesser effect, RNA-binding ability. Overall, these data confirm that CBX7 functionally interacts with H3K27me and RNA to repress the *INK4b-ARF-INK4a* locus and control senescence.

Perspectives

Some argue that lncRNAs function as transcriptional noise, having only stochastic relationships with chromatin proteins, whereas others favor the idea that lncRNAs are able to discretely collaborate with PRC1 and PRC2 complexes in regulating gene expression through H3K27me3 and H2A119ub1. Recently, *HOTAIR* and *ANRIL* have been identified as selectively binding with PRC1 and PRC2 complexes to execute histone modifications at specific loci, which strongly supports the idea that lncRNAs may function as ideal regulators for epigenetic transcriptional repression. Evidence that certain lncRNAs are expressed as monoallelic transcripts provides the ideal opportunity to distinguish paternal from maternal alleles or *vice versa*. Dosage of lncRNAs may also

provide the fine-tuning needed to restrict loading of RNA polymerase II to TSS. For the most part, transcription of lncRNAs is done by RNA polymerase II, which is strongly associated with H3K4me at promoter regions. Ironically, H3K4me is a histone activation mark controlled by *Trithorax* complex, which is composed of MLL, WDR5, MEN1, and ASH1/2. Important questions remain about the role of lncRNAs in instructing chromatin modifications. For example, how are cells able to control these bivalent histone modifications (activation marks and repression marks) in the nearby region? Are lncRNAs a consequence of random transcription by RNA polymerase II or discrete operators of chromatin function? If it is an operational process, what is the underlying mechanism? Further studies are required to determine how the PRC complexes, *Trx* complexes, and lncRNAs collaborate to regulate gene activation and/or repression to help dispel the myths behind the ghosts.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

This review article has been financially supported through awards 1R01CA154809 and 5RC1DA028776 to both M.J. Wals and M.-M. Zhou from the NIH. F. Aguilo is supported by the Ellison Medical Foundation through a Senior Scholar Award in Aging (AG-SS-2482-10) to M.J. Walsh.

Received December 20, 2010; revised April 29, 2011; accepted May 11, 2011; published OnlineFirst August 9, 2011.

References

- Schwartz YB, Pirrotta V. Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet* 2007;8:9–22.
- Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, et al. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 2004;431:873–8.
- Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* 2006;26:2560–9.
- Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 2002;298:1039–43.
- Min J, Zhang Y, Xu RM. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* 2003;17:1823–8.
- Simon JA, Lange CA. Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutat Res* 2008;647:21–9.
- Jacobs JJ, Kieboom K, Marino S, DePinho RA, van Lohuizen M. The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature* 1999;397:164–8.
- Gil J, Bernard D, Martínez D, Beach D. Polycomb CBX7 has a unifying role in cellular lifespan. *Nat Cell Biol* 2004;6:67–72.
- Ringrose L, Paro R. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* 2007;134:223–32.
- Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 2009;106:11667–72.
- Pauli A, Rinn JL, Schier AF. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* 2011;12:136–49.
- Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* 2009;23:1831–42.
- Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010;464:1071–6.
- Tano K, Mizuno R, Okada T, Rakwal R, Shibato J, Masuo Y, et al. MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett* 2010;584:4575–80.
- Astuti D, Latif F, Wagner K, Gentle D, Cooper WN, Catchpoole D, et al. Epigenetic alteration at the DLK1-GTL2 imprinted domain in human neoplasia: analysis of neuroblastoma, pheochromocytoma and Wilms' tumour. *Br J Cancer* 2005;92:1574–80.
- Gil J, Peters G. Regulation of the *INK4b-ARF-INK4a* tumour suppressor locus: all for one or one for all. *Nat Rev Mol Cell Biol* 2006;7:667–77.
- Popov N, Gil J. Epigenetic regulation of the *INK4b-ARF-INK4a* locus: in sickness and in health. *Epigenetics* 2010;5:685–90.
- Bernard D, Martínez-Leal JF, Rizzo S, Martínez D, Hudson D, Visakorpi T, et al. CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the *Ink4a/Arf* locus. *Oncogene* 2005;24:5543–51.
- Maertens GN, El Messaoudi-Aubert S, Racek T, Stock JK, Nicholls J, Rodriguez-Niedenführ M, et al. Several distinct polycomb complexes regulate and co-localize on the *INK4a* tumor suppressor locus. *PLoS ONE* 2009;4:e6380.
- Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 2008;451:202–6.
- El Messaoudi-Aubert S, Nicholls J, Maertens GN, Brookes S, Bernstein E, Peters G. Role for the MOV10 RNA helicase in

- polycomb-mediated repression of the INK4a tumor suppressor. *Nat Struct Mol Biol* 2010;17:862–8.
22. Pasmant E, Laurendeau I, Héron D, Vidaud M, Vidaud D, Bièche I. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* 2007;67:3963–9.
 23. Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, et al. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 2011;30:1956–62.
 24. Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 2010;38:662–74.
 25. Jarinova O, Stewart AF, Roberts R, Wells G, Lau P, Naing T, et al. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler Thromb Vasc Biol* 2009;29:1671–7.
 26. Rodriguez C, Borgel J, Court F, Cathala G, Forné T, Piette J. CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochem Biophys Res Commun* 2010;392:129–34.
 27. Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet* 2010;6:e1001233.
 28. Pasmant E, Sabbagh A, Vidaud M, Bièche I. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J* 2011;25:444–8.
 29. Liu Y, Sanoff HK, Cho H, Burd CE, Torrice C, Mohlke KL, et al. INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS ONE* 2009;4:e5027.
 30. Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs associated with multiple disease phenotypes correlate with ANRIL expression. *PLoS Genet* 2010;6:e1000899.