Gene-specific repression of the p53 target gene PUMA via intragenic CTCF – Cohesin binding

Nathan P. Gomes and Joaquín M. Espinosa

*Genes Dev.* 2010 24: 1022-1034
Access the most recent version at doi: 10.1101/gad.1881010

Supplemental Material

http://genesdev.cshlp.org/content/suppl/2010/05/05/24.10.1022.DC1.html

References

This article cites 68 articles, 25 of which can be accessed free at:
http://genesdev.cshlp.org/content/24/10/1022.full.html#ref-list-1

Open Access

Freely available online through the Genes & Development Open Access option.

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)
Gene-specific repression of the p53 target gene PUMA via intragenic CTCF–Cohesin binding

Nathan P. Gomes and Joaquín M. Espinosa

Howard Hughes Medical Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309, USA

The p53 transcriptional program orchestrates alternative responses to stress, including cell cycle arrest and apoptosis, but the mechanism of cell fate choice upon p53 activation is not fully understood. Here we report that PUMA (p53 up-regulated modulator of apoptosis), a key mediator of p53-dependent cell death, is regulated by a noncanonical, gene-specific mechanism. Using chromatin immunoprecipitation assays, we found that the first half of the PUMA locus (~6 kb) is constitutively occupied by RNA polymerase II and general transcription factors regardless of p53 activity. Using various RNA analyses, we found that this region is constitutively transcribed to generate a long unprocessed RNA with no known coding capacity. This permissive intragenic domain is constrained by sharp chromatin boundaries, as illustrated by histone marks of active transcription (histone H3 Lys9 trimethylation [H3K4me3] and H3K9 acetylation [H3K9Ac]) that precipitously transition into repressive marks (H3K9me3). Interestingly, the insulator protein CTCF (CCCTC-binding factor) and the Cohesin complex occupy these intragenic chromatin boundaries. CTCF knockdown leads to increased basal expression of PUMA concomitant with a reduction in chromatin boundary signatures. Importantly, derepression of PUMA upon CTCF depletion occurs without p53 activation or activation of other p53 target genes. Therefore, CTCF plays a pivotal role in dampening the p53 apoptotic response by acting as a gene-specific repressor.

Keywords: p53; PUMA; CTCF; cohesin; apoptosis; noncoding RNA

Supplemental material is available at http://www.genesdev.org.

Received November 1, 2009; revised version accepted March 16, 2010.

The cellular response to stress involves a plethora of interconnecting sensor and effector pathways [Abraham 2001]. Of critical importance to the stress response is the tumor suppressor p53 network, which mediates alternative cellular fates, including cell cycle arrest, senescence, or apoptosis [Vousden and Prives 2009]. p53 functions primarily as a transcription factor by binding to cognate p53-responsive elements [p53REs] in downstream target genes, leading to their transcriptional activation or repression [Laptenko and Prives 2006]. Induction of p53-mediated cell cycle arrest involves the transcriptional activation of the cyclin-dependent kinase inhibitor p21 [CDKN1A] and the inhibitor of G2/M transition, SFN [14-3-3σ] [Harper et al. 1993; Yang et al. 2006]. p53-dependent apoptosis involves the expression of genes in the mitochondrial apoptotic pathway, such as p53 up-regulated modulator of apoptosis [PUMA [BBC3]], APAF1, and BAX [Miyashita and Reed 1995; Zou et al. 1997; Nakano and Vousden 2001; Yu et al. 2001], and in the death receptor pathway, such as FAS and DR5 [TNFRSF10B] [Muller et al. 1998; Takimoto and El-Deiry 2000]. Of note, the p53 transcriptional program is remarkably flexible, with stimulus- and cell type-specific variations being relatively common [Zhao et al. 2000; Espinosa 2008]. This regulatory diversity allows cells to mount alternative responses to stress in a context-dependent manner. Understanding the molecular mechanisms driving this pleiotropy is of much clinical relevance, as this knowledge may lead to better therapeutic strategies harnessing the apoptotic potential of p53 for selective elimination of cancer cells. In this context, deciphering gene-specific regulatory mechanisms within the p53 network is of critical importance, and a wealth of research has produced a number of working models. At one end of the spectrum, p53-centric models postulate that p53 itself orchestrates cell fate choice. This would be achieved by p53 post-translational modifications or p53-binding proteins capable of modulating the transcriptional activity of p53 in a gene-specific manner. For example, the ability of p53 to transactivate proapoptotic genes can be enhanced by phosphorylation of Ser46, acetylation of Lys120, or association with the ASPP
cofactor (Oda et al. 2000; Samuels-Lev et al. 2001; Sykes et al. 2006; Tang et al. 2006). At the other end of the spectrum, contextual models postulate that p53 transactivation potential is fairly invariant, and that the p53-dependent transcriptome is strongly influenced by the different contexts in which p53 acts (e.g., chromatin landscape and target core promoter elements). For example, hCAS/CSE11 promotes p53-dependent transcription of specific apoptotic genes by a p53-autonomous mechanism likely involving histone demethylation (Tanaka et al. 2007). Along these lines, we demonstrated that cell cycle arrest genes within the p53 network [e.g., p21] harbor large amounts of preloaded RNA polymerase II [RNAPII] on their core promoters, and that their expression is regulated at post-initiation and elongation stages of the transcriptional cycle [Espinosa et al. 2003; Gomes et al. 2006]. In contrast, proapoptotic genes carry much lower levels of preloaded RNAPII, which may be indicative of a different mode of regulation as compared with cell cycle arrest genes. Recently, it was demonstrated that the core promoter elements found at p53 target genes determine the rates of both of preinitiation complex (PIC) assembly and RNAPII reinitiation in a p53-autonomous fashion (Morachis et al. 2010).

Here we demonstrate the existence of a novel gene-specific mechanism regulating the expression of the potent proapoptotic p53 target gene PUMA. Unexpectedly, we found that, although PUMA harbors little preloaded RNAPII at its two previously described alternative promoters, it nonetheless undergoes constitutive transcription throughout the first 6 kb of the intragenic region, with levels of intragenic RNAPII being as high as those found on the core promoters of cell cycle arrest genes. These intragenic RNAPII complexes are constitutively phosphorylated on the RNAPII C-terminal domain (CTD) at Ser5 and Ser2, indicative of actively elongating complexes. Several general transcription factors (GTFs) [e.g., TBP, TFIIB, and TFIIF], components of the Mediator complex, and elongation factors [e.g., Positive Transcription Elongation Factor b [P-TEFb]] also accumulate constitutively throughout this region. Accordingly, we demonstrate that the first half of the PUMA locus undergoes constitutive transcription, which gives rise to an unprocessed non-coding RNA species. We found that the PUMA locus harbors a distinct intragenic chromatin architecture wherein histone marks indicative of active transcription [histone H3 Lys9 trimethylation [H3K4me3] and H3K9 acetylation [H3K9Ac]] are constrained within a 6-kb region, with flanking regions harboring the repressive mark H3K9me3. Importantly, noncanonical intragenic occupancy of CCCTC-binding factor (CTCF) and Cohesin complexes define these chromatin boundaries. CTCF knockdown leads to an increase in basal PUMA mRNA and protein levels. These results demonstrate that CTCF mediates a novel p53-autonomous mechanism regulating the basal expression of a potent proapoptotic p53 target gene by maintaining intragenic chromatin boundaries.

Results

Differential RNAPII core promoter occupancy among p53 target genes

Recent data demonstrate that up to 70% of human genes have RNAPII associated with their proximal promoters, regardless of their activation status [Guenther et al. 2007]. Many of the genes that fall under this category are associated with stress responses and developmental programs. Among genes within the tumor suppressor p53 transcriptional network, those involved in cell cycle arrest and DNA repair display markedly different proximal promoter RNAPII occupancy profiles when compared with those involved in apoptosis [Espinosa et al. 2003]. In untreated HCT116 cells, the p21, SFN, and GADD45a genes [cell cycle arrest] harbor far more total RNAPII at their core promoters [P] when compared with the proapoptotic genes FAS, APAF1, and PUMA [Fig. 1A]. Upstream control regions [C] for each gene are shown. Of note, PUMA has two transcriptional start sites [see Fig. 2A; Supplemental Fig. 1], both of which show little preloaded RNAPII. Upon transcriptional activation of these genes with the anti-metabolite 5-fluorouracil [5-FU], the total amount of RNAPII associated with all proximal promoters increases, but cycle arrest genes still harbor significantly more RNAPII than proapoptotic promoters. RNAPII CTD phosphorylation of Ser5 and Ser2 [SSP and S2P] is generally thought to be involved in post-RNAPII recruitment stages of the transcription cycle [Sims et al. 2004]. Among p53 target gene promoters, SSP patterned mimics that of total RNAPII, where cell cycle arrest genes harbor far more SSP than proapoptotic genes [Espinosa et al. 2003]. Upon 5-FU treatment, SSP signals are induced across all genes to a greater extent than total RNAPII levels, suggesting that a fraction of preloaded RNAPII exists in a hypophosphorylated state prior to activation [Fig. 1A]. We demonstrated previously that genes involved in cell cycle arrest [e.g., p21] are regulated at post-initiation/elongation steps of the transcription cycle [Gomes et al. 2006; Donner et al. 2007]. This led us to examine how genes involved in the apoptotic response are transcriptionally regulated.

PUMA: p53-dependent activation and apoptotic function

The PUMA gene is comprised of an ~12-kb locus, which harbors two transcription start sites and encodes four known RNA species: PUMAα, PUMAβ, PUMAδ, and PUMAγ [Supplemental Fig. 1]. Previous work suggests that PUMAβ is constitutively transcribed and PUMAα is p53-inducible, whereas the regulation and functions of PUMAδ and PUMAγ remain more elusive [Vosudden 2005; Tampio et al. 2009]. PUMA’s known roles in the apoptotic pathway are to antagonize the function of prosurvival members of the Bcl-2 family and facilitate Bax translocation to the mitochondria [Yee and Vosudden 2008; Zhang et al. 2009]. Despite being considered one of the most important proapoptotic p53 target genes, little is known about the mechanisms regulating its expression.
Treatment of HCT116 cells with 5-FU leads to the rapid accumulation of the p53 protein with a concomitant accumulation of PUMAα (Fig. 1B). PUMAα accumulation is dependent on p53, as demonstrated by 5-FU treatment in isogenic p53-null HCT116 cells (Fig. 1B; Yu et al. 2001). Additionally, the PUMAα protein does not accumulate in isogenic PUMA-null HCT116 cells (Fig. 1B; Yu et al. 2003). Real-time RT–PCR (Q-RT–PCR) analysis using exon 4-specific primers demonstrates that total PUMA mRNA accumulates over time in a p53-dependent manner (Fig. 1C; Nakano and Vousden 2001; Yu et al. 2001). Given PUMA’s importance within the p53-mediated apoptotic response, we chose it as a model to study the transcriptional regulation of proapoptotic genes.

The PUMA locus constitutively harbors RNAPII within half of the gene body

In order to study the mechanisms of p53-dependent transcriptional activation of PUMA, we developed a high-resolution chromatin immunoprecipitation (ChIP)

![Figure 1](genesdev.cshlp.org) Differential RNAPII preloading on p53 target gene promoters and the role of PUMA in p53-dependent apoptosis. [A] ChIP analysis in HCT116 cells of p53 target promoters before and after 8 h of 5-FU treatment (375 μM) with antibodies recognizing total RNAPII (RNAPII) and the phosphorylated isoforms of Ser5 of the RNAPII CTD (S5P). Values are presented as the percentage of maximum ChIP signal. ChIP-enriched DNA was quantified by real-time PCR analysis using amplicons surrounding the core promoters (P) and upstream control regions (C) for each loci. Note that PUMA harbors two core promoters, denoted p60 (upstream promoter) and α (downstream promoter). [B] Western blot analysis demonstrating p53 and PUMAα accumulation in response to 5-FU treatment in isogenic HCT116 p53+/+, p53-/- and PUMA-/- cells. Nucleolin serves as a loading control. [C] Q-RT–PCR analysis of PUMA mRNA accumulation in response to 5-FU treatment. All values are normalized to 18S ribosomal RNA levels and are shown as fold induction over control HCT116 p53+/+ cells. [D] Apoptotic index assays of cells treated with 5-FU for the designated times, stained with annexin V and propidium iodide (PI), and analyzed by flow cytometry. Histograms represent the percentage of cells positive for annexin V staining. Empty vector serves as a control for PUMA overexpression.

![Figure 2](genesdev.cshlp.org) RNAPII patterning on the PUMA and p21 loci under basal and p53-activated states. [A] Linear scale maps of the PUMA and p21 loci showing the locations of p53REs, transcription start sites (designated by arrows), exon/intron structure, and polyadenylation signal [AATAAA]. The location of 20 real-time PCR amplicons used in ChIP assays are also shown. The numbers represent the position of the center of each amplicon relative to the transcription start sites (the more 5’ transcription start site for PUMA). ChIP assays were performed with whole-cell extracts from control and 5-FU-treated (8 h) HCT116 p53+/+ cells with antibodies recognizing total p53 (p53) ([B]), total RNAPII (RNAPII) ([C]), phosphorylated Ser5 CTD of RNAPII (S5P) ([D]), and phosphorylated Ser2 CTD of RNAPII (S2P) ([E]). ChIP-enriched DNA was quantified by real-time PCR. Values are presented as the percentage of maximum ChIP signal. The number of biological ChIP and PCR replicates for each epitope is indicated. The locus maps are a condensed scale model of those seen in A, and asterisks represent the p53REs. The gray bands represent the transcribed region of each locus.

Gomes and Espinosa

Cold Spring Harbor Laboratory Press
assay, allowing us to observe spatial and temporal protein occupancy on the PUMA locus. From our previous work [Espinosa et al. 2003; Gomes et al. 2006; Donner et al. 2007], we have a good understanding of p53-regulated transcription of the p21 locus, and it is therefore used for comparative purposes. Figure 2A shows a scale representation of the two loci. As expected, chromatin-associated p53 follows total cellular p53 levels both before and after 8 h of 5-FU treatment [Fig. 2B]. On the PUMA locus, p53 associates with its two known moderate-affinity, clustered intronic p53REs [amplicons +1313] [Nakano and Vousden 2001; Yu et al. 2001], while on the p21 locus, p53 associates with its known high-affinity REs located upstream of the core promoter [amplicons −2283 and −1391, respectively] [Espinosa and Emerson 2001]. Note that for all ChIP data, the PUMA and p21 loci are on the same scale relative to maximum ChIP signal. Consistent with the observation that p53 target genes harbor a broad range of binding site affinities [Szak et al. 2001], we observe that the p21 locus harbors approximately fivefold more p53 than the PUMA locus in an induced state. RNAPII ChIP analysis of the PUMA locus lead to the very surprising and unprecedented observation that RNAPII associates throughout the first 6 kb of PUMA, with a peak of accumulation within intron 3 [amplicon +6197], at a level greater than that observed at the preloaded p21 promoter [amplicon −20] [Fig. 2C]. Upon 5-FU treatment, a small but reproducible drop is seen in the accumulated RNAPII in intron 3 of PUMA, concurrent with an increase at the 5′ end [approximately threefold] and 3′ region of the gene [approximately fivefold] [Fig. 2C, see inset]. RNAPII occupancy on the p21 locus follows a more “canonical” pattern, with accumulation only around the proximal promoter before activation, while upon 5-FU stimulation, polymerase increases modestly around the start site [less than twofold] and dramatically in the body of the gene [−15-fold to 20-fold]. This unexpected accumulation of RNAPII in the body of the PUMA locus is largely p53-independent, as demonstrated by similar patterning in HCT116 p53-null cells (Supplemental Fig. 2A). Additionally, this patterning of RNAPII is not cell type-specific, as it is also observed in U2OS and BV173 cell lines (Supplemental Fig. 2B). Moreover, this patterning of RNAPII has also been observed in genome-wide RNAPII ChIP sequencing [ChIP-seq] from CD4+ T cells [Supplemental Fig. 2C; Barski et al. 2007]. Understanding what this curious RNAPII patterning means to the expression of PUMA dictated the direction of our study.

Phosphorylation of the RNAPII CTD heptads is associated with post-initiation and elongation stages of the transcription cycle, and can be indicative of RNAPII behavior [e.g., direction of transcription]. In order to elucidate the nature of the total RNAPII patterning observed on PUMA, we next investigated the patterning of SS2 and S2P. Basal SS2 levels on PUMA mimics total polymerase, associating with the first 6 kb of the transcribed region peaking within intron 3 [amplicon +6197]. Upon 5-FU stimulation, the SS2 signal increases minimally around the peak in intron 3, and modestly [approximately threefold] in the 5′ and the 3′ ends of the gene [Fig. 2D]. S2P patterning on p21 is representative of inducible genes, with an accumulation around the transcription start site before activation [amplicon −20], and then larger accumulation upon activation at the promoter and throughout the gene body. Typically, S2P levels increase progressively toward the 3′ end of genes [Ahn et al. 2004; Sims et al. 2004]. A clear example of this occurs on the p21 locus, with low basal levels of S2P at the core promoter that increase toward the 3′ end of the gene, indicative of basal transcription, and an approximately sevenfold increase in signals upon activation [Fig. 2E]. In contrast, basal levels of S2P on the PUMA locus peak within the intragenic region along with total RNAPII [amplicon +6197], then rapidly drop off, and then modestly begin to increase toward the 3′ end of the gene. In the 5-FU-stimulated state, S2P levels for the first half of the gene mirror that seen in the uninduced state, while increased S2P levels are observed at the 3′ end. In a most parsimonious interpretation, these observations suggest that the PUMA locus undergoes constitutive basal transcription from its known promoters and pauses within the body of the gene, 6 kb downstream from the core promoters. Upon activation by p53, RNAPII is able to overcome this transcriptional pause and progress toward the end of the gene.

Dispersed and constitutive association of GTFs and elongation factors on the PUMA locus

RNAPII recruitment to a given DNA region and activation into an elongation-competent form is dependent on a myriad of GTFs, the Mediator complex, diverse chromatin-modifying and chromatin remodeling complexes, and various elongation factors [Lee and Young 2000; Smale and Kadonaga 2003; Casamassimi and Napoli 2007]. Intrigued by the curious pattern of RNAPII occupancy and CTD phosphorylation pattern within PUMA, we examined the association of many of these factors. The TATA-binding protein (TBP) exists within large multisubunit complexes involved in PIC formation (e.g., TFIIID) and associates with core promoter regions [Burley and Roeder 1996; Hochheimer and Tjian 2003]. Expectedly, TBP occupancy is restricted to the 5′ end of the p21 locus and is largely unaffected by the activation status of this preloaded gene [Fig. 3A]. Strikingly, TBP is associated on the PUMA locus throughout the first 6 kb of the intragenic region. Upon 5-FU stimulation, TBP remains throughout the first 6 kb of the locus, but the levels drop [approximately twofold to threefold] around the third exon of PUMA [amplicons +6014 and +6197]. TFIIB also associates with core promoters as an integral part of the PIC, and is involved in RNAPII recruitment and reinitiation [Deng and Roberts 2007]. On the p21 promoter, TFIIB association increases significantly upon gene activation [Fig. 3B]. As seen for TBP, TFIIB is observed throughout the first 6 kb of the PUMA locus prior to activation, but is found mostly around the known transcription start sites [amplicons +8 and +1313] and the peak sites of RNAPII accumulation [amplicons +5686, +6014, and +6197]. Upon activation, TFIIB levels increase around the known
transcription start sites, while remaining constant in the middle of the gene. On the p21 locus, TFIIIF associates basally with the core promoter, increases upon stimulation, and spreads into the first third of the gene body. This is consistent with TFIIIF's role in proper PIC positioning and facilitating promoter escape (Conaway et al. 2000). As with the other GTFs, TFIIIF is associated throughout the first 6 kb of the PUMA locus before 5-FU treatment, with peaks of association around the known transcription start sites [amplicons +8 and +1313] and the middle of the gene [amplicon +6197]. Upon stimulation, TFIIIF increases only slightly (approximately twofold) around the known transcription start sites, while remaining constant throughout the rest of the gene. Taken together, these data demonstrate that the first half of the PUMA locus harbors marks of PIC formation, with major sites of accumulation around the known core promoters and 6 kb within the gene body.

The Mediator complex is a large multisubunit complex that allows communication between DNA-binding proteins and RNAPII, and has been demonstrated to play roles in multiple steps of the transcription cycle (Casamassimi and Napoli 2007). On p21, other p53-inducible genes, and serum responsive genes, the CDK8 subunit of Mediator is recruited to the core promoter upon activation (Fig. 3D; Donner et al. 2007, 2010). On the PUMA gene, CDK8 is constitutively associated with the known transcription start sites [amplicons +8 and +1313] and the intragenic site of RNAPII accumulation [amplicon +6197]. Upon activation, CDK8 levels increase very minimally around the PUMA core promoter, while remaining constant throughout the rest of the gene [Fig. 3D]. The Negative Elongation Factor (NELF) complex has been demonstrated to promote RNAPII pausing around core promoter regions (Yamaguchi et al. 1999; Gilchrist et al. 2008). On p21, NELF associates solely with the core promoter, and its levels remain unchanged in an activated state [Fig. 3E]. As with the GTFs, and in a pattern similar to that of RNAPII, NELF associates with the first 6 kb of the PUMA locus, with a peak around the middle of the gene [amplicon +6197]. Upon activation, NELF levels decrease slightly around the middle of the gene, but the overall pattern remains unchanged [Fig. 3E]. As its name implies, the P-TEFb complex promotes functional elongation and overcomes the effects of negative elongation factors. On inducible genes such as p21, basal levels of P-TEFb (CDK9) are relatively low, with a small accumulation around the core promoter (Fig. 3F). Upon activation, levels increase dramatically around the transcription start sites as well as the body of the gene. On PUMA, P-TEFb is again associated with the first half of the locus, with peaks of association around the core promoter and exon 3 [Fig. 3F]. Upon 5-FU stimulation, P-TEFb association increases minimally throughout the first half of the gene body. Taken together, these data suggest that the first 6 kb of the PUMA locus is undergoing active transcription elongation. Overall, the broad association of GTFs, Mediator, and elongation factors within the first 6 kb of the locus begs the question: What types of RNA species are being produced from the PUMA locus before and during activation?

The 5’ half of the PUMA locus is constitutively transcribed

To investigate the RNA species encoded by the PUMA locus under basal and activated states, we analyzed total RNA from control and 5-FU-treated HCT116 cells on Affymetrix GeneChip Human Exon 1.0 ST arrays. These arrays measure sense strand RNA abundance for all annotated exons, using multiple probes per exon. As expected, all exons of the p21 locus [Fig. 4A] as well as other p53 targets [Supplemental Fig. 3A] display equivalent induction upon 5-FU stimulation. Surprisingly, on PUMA, exons 1a, 2, and 3 show modest induction [<1.5-fold], while exons 1b and 4 are more strongly induced upon 5-FU stimulation (1.8-fold and 4.6-fold, respectively). This result demonstrates differential expression levels [nonstoichiometric] of PUMA exons, and suggests that the PUMA locus is constitutively transcribed in the sense direction from exon 1a to exon 3. Furthermore, it also suggests that exon 1b is also responsive to p53 activation. Of note, we have yet to identify any other p53 target gene that displays a similar
Intron 3 PUMA RNA RACE. A diagram of the strand-specific exon 4 and intron 3 probes as described in strand-specific Northern blots of total and polyA-intron 3, whose relative positions are shown on the PUMA locus representing exons 1b, 1a, 2, 3, and 4, as well as the first 100 bp of intron 3 (where we observed a large accumulation of RNAPII by ChIP) using end-labeled 80-mer single-stranded oligonucleotides. As expected, all exonic probes detected accumulation of a 5-FU-inducible RNA species corresponding to the known PUMA mRNA (Fig. 4C, PUMA). To our surprise, but very consistent with the exon array data, for exons 1a, 2, and 3, as well as intron 3, we observed a larger, constitutively present RNA species. Following recommendations by the Encyclopedia of DNA Elements [ENCODE] consortium, we dubbed this novel RNA species as PUMA-TUF [transcript of unknown function] (Fig. 4C). Of note, reverse complementary probes did not detect any of these RNA species, thus discouraging the notion of significant antisense transcription taking place [data not shown; see the Discussion]. In order to help define the nature of PUMA-TUF, we enriched for polyadenylated [polyA+] RNAs from control and 5-FU-stimulated cells and subjected them to sense strand-specific Northern blots. As seen in Figure 4D, polyA+ purification significantly enriches the known PUMA mRNA (exon 4 probe) relative to total RNA. Interestingly, polyA+ enrichment also leads to a loss of PUMA-TUF relative to total RNA (intron 3 probe). In sum, these data demonstrate the existence of a large, unprocessed, and constitutively transcribed RNA species containing PUMA exons 1a, 2, and 3, and part of intron 3. In order to define the 3′ end of PUMA-TUF, we performed rapid amplification of cDNA ends assays [3′-RACE] with a fixed forward primer designed within the last 25 bp of exon 3 and four alternative reverse primers targeting intron 3 in increments of 100 bp. All of these primer sets amplify in the same linear range when using gDNA as a template [Fig. 4E]. In contrast, when cDNA from untreated HCT116 cells is used as a template, we observe amplification of only the first 100 bp of intron 3. This suggests that PUMA-TUF ends between 100 and 200 bp downstream from exon 3. In agreement with our observations, the Affymetrix Transcriptome Project detected a nuclear nonpolyadenylated RNA encoded by the first 6 kb of the PUMA locus in HepG2 cells, while cytoplasmic polyA+ RNA from the same cells contains only the exons [Supplemental Fig. 3B, Cheng et al. 2005]. Overall, these results indicate constitutive sense transcription of the PUMA locus, starting around exon 1a and terminating within the first 200 bp of the third intron, to produce a large unprocessed RNA, which is possibly retained in the nucleus. Of note, our analyses failed to detect any known sequences of microRNAs, small nuclear RNAs, or ORFs giving rise to known proteins, suggesting that PUMA-TUF is a noncoding RNA [data not shown].

Figure 4. Analysis of RNA species encoded by the PUMA locus. [A] Exon-level sense strand RNA analysis using Affymetrix exon arrays with total RNA from control and 5-FU-treated (12 h) HCT116 cells. The results are an average of two independent experiments. Values are expressed as the normalized fold induction for each individual exon. [B] Semiquantitative RT–PCR (top) and Q-RT–PCR (bottom) of PUMA-derived RNAs containing exons 3 and 4. Semiquantitative RT–PCR analysis is relative to a gDNA standard. Q-RT–PCR analysis was quantified as described for Figure 1C. [C] Sense strand-specific Northern blots. Total RNA from control and 5-FU-treated [16 h] samples were probed with 80-mer strand-specific oligonucleotides representing exons 1b, 1a, 2, 3, and 4, as well as the first 100 bp of intron 3, whose relative positions are shown on the PUMA locus map. 185 ribosomal RNA serves as a loading control. [D] Sense strand-specific Northern blots of total and polyA+ -enriched RNA. Equal masses of each preparation were probed with sense strand-specific exon 4 and intron 3 probes as described in C. [E] Intron 3 PUMA RNA RACE. A diagram of the PUMA exon 3/intron 3 boundary showing the location of a fixed forward primer and alternative reverse primers at 100-bp increments. PCR was performed with gDNA and cDNA from untreated HCT116 p53−/− cells.

pattern of differential exon level expression. In order to confirm these observations, we performed RT–PCR assays to measure the relative levels of exons 3 and 4. Using Q-RT–PCR, we observed that exon 4 is induced more than ninefold upon 5-FU stimulation, while exon 3 is induced approximately threefold [Fig. 4B, bottom]. Using semiquantitative RT–PCR with a fixed genomic DNA [gDNA] standard, we found that significant amounts of exon 3 can be detected before p53 activation with only a modest induction after 5-FU treatment [Fig. 4B, top]. In stark contrast, there are undetectable levels of exon 4 before 5-FU treatment, with a large increase following it, again consistent with the nonstoichiometric levels of exons 1a, 2, and 3 relative to exons 1b and 4 observed in the exon arrays. In order to further characterize the RNA species being expressed from the PUMA locus, we performed sense strand-specific Northern blots to each exon of PUMA as well as the first 100 base pairs [bp] of intron 3 (where we observed a large accumulation of RNAPII by ChIP) using end-labeled 80-mer single-stranded oligonucleotides. As expected, all exonic probes detected accumulation of a 5-FU-inducible RNA species corresponding to the known PUMA mRNA (Fig. 4C, PUMA). To our surprise, but very consistent with the exon array data, for exons 1a, 2, and 3, as well as intron 3, we observed a larger, constitutively present RNA species. Following recommendations by the Encyclopedia of DNA Elements [ENCODE] consortium, we dubbed this novel RNA species as PUMA-TUF [transcript of unknown function] (Fig. 4C). Of note, reverse complementary probes did not detect any of these RNA species, thus discouraging the notion of significant antisense transcription taking place [data not shown; see the Discussion]. In order to help define the nature of PUMA-TUF, we enriched for polyadenylated [polyA+] RNAs from control and 5-FU-stimulated cells and subjected them to sense strand-specific Northern blots. As seen in Figure 4D, polyA+ purification significantly enriches the known PUMA mRNA (exon 4 probe) relative to total RNA. Interestingly, polyA+ enrichment also leads to a loss of PUMA-TUF relative to total RNA (intron 3 probe). In sum, these data demonstrate the existence of a large, unprocessed, and constitutively transcribed RNA species containing PUMA exons 1a, 2, and 3, and part of intron 3. In order to define the 3′ end of PUMA-TUF, we performed rapid amplification of cDNA ends assays [3′-RACE] with a fixed forward primer designed within the last 25 bp of exon 3 and four alternative reverse primers targeting intron 3 in increments of 100 bp. All of these primer sets amplify in the same linear range when using gDNA as a template [Fig. 4E]. In contrast, when cDNA from untreated HCT116 cells is used as a template, we observe amplification of only the first 100 bp of intron 3. This suggests that PUMA-TUF ends between 100 and 200 bp downstream from exon 3. In agreement with our observations, the Affymetrix Transcriptome Project detected a nuclear nonpolyadenylated RNA encoded by the first 6 kb of the PUMA locus in HepG2 cells, while cytoplasmic polyA+ RNA from the same cells contains only the exons [Supplemental Fig. 3B, Cheng et al. 2005]. Overall, these results indicate constitutive sense transcription of the PUMA locus, starting around exon 1a and terminating within the first 200 bp of the third intron, to produce a large unprocessed RNA, which is possibly retained in the nucleus. Of note, our analyses failed to detect any known sequences of microRNAs, small nuclear RNAs, or ORFs giving rise to known proteins, suggesting that PUMA-TUF is a noncoding RNA [data not shown].
A well-defined intragenic chromatin boundary exists within the PUMA locus

While it is quite surprising that the first half of the PUMA locus harbors RNAPII and associated factors and undergoes constitutive transcription, it is equally surprising that these activities sharply cease in the last half of the gene. This led us to investigate whether the chromatin architecture of the PUMA locus was responsible for constraining this permissive basal transcription. The total histone H3 density across PUMA is relatively constant [Fig. 5A], suggesting that constitutive transcription on PUMA is not due to a gross depletion of nucleosomes. Next, we investigated known marks of active and repressed chromatin. H3K4me3 defines the 5′ end of transcriptionally competent genes (Shilatifard 2008). Genome-wide data for H3K4me3 shows an average peak at 500 bp to 1 kb downstream from transcriptional start sites while being absent from the rest of the gene body, and this canonical pattern is clearly observed on p21 (Fig. 5B, Guenther et al. 2007). Notably, on the PUMA locus, the H3K4me3 signal is seen throughout the first 6 kb of the locus, with signals dropping off precipitously within intron 3 (amplicon +7021). Another mark of active transcription at enhancers and core promoters is H3K9Ac [Agalioti et al. 2002]. Under basal conditions, there are peaks of H3K9Ac observed just downstream from the core promoter (amplicons +182 and +507) and around the distal p53-binding sites (amplicons −2283 through −1391) on the p21 locus (Fig. 5C). Upon activation, these levels increase in both regions. Strikingly, on the PUMA locus, H3K9Ac is found throughout the first 6 kb of the locus, with a peak in the middle of the gene (amplicon +6197) [Fig. 5C]. Upon 5-FU stimulation, H3K9Ac increases around the intronic p53-binding sites (amplicon +1313) while remaining unchanged throughout the rest of the gene. Therefore, the PUMA locus harbors constitutive histone marks of transcriptional activation constrained within the first half of the intragenic region. On the other hand, specific histone marks are also associated with repressed chromatin. For example, H3K9me3 is often associated with heterochromatin and a repressed transcriptional state [Stewart et al. 2005]. Interestingly, we observed a clear peak in H3K9me3 right downstream from the drop in H3K9Ac [Fig. 5D; amplicon +7021]. Of note, H3K9me3 levels do not change significantly upon 5-FU stimulation. Taken together, these observations suggest that the PUMA locus harbors two drastically different chromatin environments, with the first 6 kb being defined by marks of “open” chromatin and active transcription followed by a region of “closed,” transcriptionally repressed chromatin. In agreement with these observations, a genome-wide DNaseI hypersensitivity study demonstrated that the first half of the PUMA locus is DNaseI-accessible (“open”), while the second half is DNaseI-resistant (“closed”) [Supplemental Fig. 4; Boyle et al. 2008].

CTCF mediates the recruitment of the Cohesin complex to define an intragenic chromatin boundary and repress basal PUMA expression

One factor commonly associated with chromatin boundaries is the zinc finger protein CTCF (Phillips and Corces 2009). Genome-wide analysis of CTCF binding demonstrated that CTCF preferentially occupies intergenic regions [Kim et al. 2007; Cuddapah et al. 2009]. To our great surprise, ChIP analysis demonstrated two peaks of intragenic CTCF association within PUMA, close to each end of the “open” chromatin domain [Fig. 6A, see peaks at amplicons +2941 and +7021]. Upon 5-FU stimulation, CTCF occupancy is reduced by ~2.5-fold in the 5′ end of the gene, while remaining constant in the middle of the gene. We also observed a large single peak of intragenic CTCF association downstream from the p21 transcriptional start site (amplicon +1775), the levels of which remain notably unchanged under activation conditions. These unusual patterns of CTCF binding were confirmed with previous genome-wide ChIP-seq studies, which demonstrate dual positions of CTCF occupancy within PUMA and single peaks of CTCF occupancy within p21 and other p53 target genes immediately downstream from their transcriptional start sites (Supplemental Fig. 5A,B, Barski et al. 2007). The presence of CTCF within PUMA prompted us to test whether any of its known interactors were also bound to this region. In fact, we found that subunits of the Cohesin complex, which facilitates CTCF insulator function [Parelho et al. 2008; Rubio et al. 2008; Wendt et al. 2008], also occupy the intragenic domain within PUMA. The Structural Maintenance of Chromosomes 1 (SMC1) and Rad21 proteins display markedly overlapping patterns of occupancy on the PUMA locus, with broad modest association around...
the transcriptional start sites and larger accumulations within the middle of the gene (amplicons +6014 and +6197) (Fig. 6B,C). Importantly, the Cohesin complex remains bound under 5-FU-stimulated conditions on PUMA. Interestingly, we also observed binding of SMC1 and Rad21 on p21 (single peak at amplicon +1775); however, the levels of cohesins decrease drastically upon transcriptional activation (Fig. 6A–C). Taken together, these data demonstrate noncanonical CTCF and Cohesin complex occupancy within PUMA and p21, and suggest a novel role for these factors in the regulation of p53 target gene expression.

To test whether CTCF has a functional role in maintaining the chromatin boundaries observed within the PUMA locus, or the regulation of other p53 target genes, we introduced control and CTCF targeting shRNAs via lentiviral transduction into HCT116 cells. As shown in Figure 6D, loss of CTCF protein levels leads to a dramatic increase in the basal levels of PUMAα but not p21 protein levels. Importantly, this increase in PUMAα is not due to stabilization of p53. The specific increase in PUMA expression in response to CTCF loss was confirmed with an independent, nonoverlapping shRNA also targeting CTCF (Supplemental Fig. 5C). Q-RT–PCR analysis showed that an 80% reduction in cellular CTCF mRNA levels leads to an ~2.5-fold increase in mature PUMA mRNA, as measured by exon 4 levels (Fig. 6E). Most interestingly, the effects of CTCF knockdown appear specific to the PUMA locus, as the expression of p21 and other p53 target genes remains unchanged. Sense strand-specific Northern blots of total RNA from control and CTCF knockdown cells demonstrate that there is a concomitant increase in PUMA exon 4 levels with a decrease in PUMA-TUF levels (Supplemental Fig. 5D). Collectively, these data suggest the function of intragenic CTCF occupancy on the PUMA locus is to reduce the basal expression levels of mature PUMA mRNA.

In order to test for a role of CTCF in the chromatin architecture of the PUMA locus, we performed ChIP analysis in CTCF knockdown cells. First, ChIP assays demonstrate a dramatic reduction (~80%) in chromatin-associated CTCF on the PUMA and p21 loci upon CTCF depletion (Fig. 7A). Second, SMC1 and Rad21 levels are also reduced on PUMA and p21 under CTCF knockdown

Figure 6. Intragenic CTCF and Cohesin occupancy regulate basal PUMA locus expression. (A–C) ChIP assays of CTCF, SMC1, and Rad21 occupancy on the PUMA and p21 loci were performed as in Figure 2. The number of ChIP and PCR replicates for each epitope is shown. (D) Western blot analysis demonstrating CTCF, PUMAα, p21, and total p53 levels in HCT116 cells expressing control and CTCF targeting shRNAs. Nucleolin serves as a loading control. (E) Q-RT–PCR analysis of CTCF and the p53 target genes PUMA [exon 4], BTG2, HDM2, FAS, PIG3, SFN, GADD45a [G45], and p21 mRNA accumulation in response to control and CTCF knockdown. All values are normalized to 18S ribosomal RNA levels and are shown as fold induction over control (CTRL) values.

Figure 7. Intragenic CTCF and Cohesin occupancy defines the chromatin state on the PUMA locus. ChIP assays of CTCF, SMC1, Rad21, total RNAPII [RNAPII], SSP, and H3K9me3 in HCT116 cells expressing control versus CTCF targeting shRNAs. The number of biological ChIP replicates and PCRs for each epitope is indicated.
conditions, but to a lesser extent (~50%) [Fig. 7B,C]. Curiously, while an increase in basal PUMA mRNA levels occurs, we observe only a small reduction in total RNAPII around the middle of the gene (amplicon +6197); however, a notable increase in RNAPII at the 3’ end of the gene is observed [Fig. 7D, inset]. Additionally, we observe a 50% reduction in the SSPE levels in the first half of the PUMA locus, with a modest increase in the second half of the locus, suggestive of transcriptional readthrough across the chromatin boundary [Fig. 7E]. On p21, RNAPII and SSPE levels remain unaltered under control and CTCF knockdown conditions, consistent with unchanged levels of p21 protein and mRNA. Interestingly, H3K9me3 levels, which define the intragenic chromatin boundary within PUMA, drop dramatically upon CTCF knockdown [Fig. 7F]. The much lower relative levels of H4K9me3 on the p21 locus are reduced as well. These data suggest that CTCF regulates an intragenic chromatin boundary within PUMA, constraining a highly transcriptionally active region, and is ultimately responsible for reducing basal expression of the protein-coding PUMA mRNA.

Discussion

The p53 network mediates cellular responses to diverse forms of stress (e.g., DNA damage, oncogene activation, hypoxia). Functioning primarily as a transcription factor, p53 regulates the expression of genes involved in cell cycle arrest, DNA repair, senescence, and apoptosis. Several studies have demonstrated the existence of multiple mechanisms that facilitate the expression of select subsets of p53 target genes in a context-dependent manner [Vousden 2006; Espinosa 2008]. Elucidating these gene-specific regulatory mechanisms will bring us closer to the design of p53-based therapies for cancer treatment. Here we demonstrate the existence of a novel gene-specific mechanism regulating the p53 target and pro-apoptotic gene PUMA. Our findings are summarized as follows: [1] A large fraction of the PUMA locus is constitutively transcribed regardless of p53 status. [2] This constitutive transcription generates a long unprocessed RNA with no known coding capacity. [3] Constitutive transcription of PUMA is constrained by an intragenic chromatin boundary defined by intragenic CTCF and Cohesin complex binding. [4] Upon p53 activation, RNAPII overcomes this intragenic boundary, resulting in functional PUMA mRNA accumulation. [5] CTCF knockdown derepresses basal PUMA transcription without p53 activation, thus acting as a gene-specific repressor within this network.

Recent studies using genome-wide RNA analyses have generated an updated and enigmatic view of how eukaryotic genomes are organized and transcribed [Kapranov et al. 2007]. These studies generated several important conclusions. First, the portion of the genome that is actually transcribed is much higher than previously expected. The ENCODE consortium estimates that up to 93% of the human genome may be transcribed [Lindberg and Lundeberg 2010]. Second, much of this RNA remains in the nucleus. It has been calculated that nuclear RNA exceeds in complexity the cytoplasmic RNA by fivefold, and that 80% of this nuclear RNA corresponds to unannotated regions of the genome. Most of these RNA species have no known coding capacity, and the ENCODE project has proposed the use of the term TUF for these RNAs [Kapranov et al. 2007]. Third, the linear model of genome organization (where functional elements are arranged linearly along the DNA and are interrupted by nonfunctional sequences) needs considerable revision. For example, a given DNA fragment can encode multiple overlapping RNA species produced by both sense and antisense transcription from intragenic and intergenic regions, with only select species having known coding capacity. Finally, another layer of complexity is provided by the existence of “long-range interconnected transcription” and “chimeric RNAs,” in which different portions of a continuous RNA species are encoded by DNA fragments separated by multiple intervening gene loci or by DNA sequences from different chromosomes [Maher et al. 2009]. It is in this context that research efforts aimed at characterizing specific gene loci displaying this degree of complexity may provide new insights into genome organization and regulation. The PUMA locus clearly falls in this category.

The surprising observation that RNAPII occupies the first 6 kb of the PUMA locus has not, to our knowledge, been previously reported in the literature for any other gene. A priori, three hypotheses could explain this unusual RNAPII patterning on PUMA: [1] The locus is constitutively transcribed from the known PUMA transcriptional start sites, pauses within the middle of the gene, and only upon p53-mediated activation does RNAPII overcome this pause to make functional PUMA mRNA. [2] The peak of polymerase in the middle of the gene does not represent an elongating complex that is paused; rather, it is polymerase initiating and traveling in the antisense direction, which could lead to transcriptional interference against PUMA mRNA production. [3] The entire first 6 kb of the PUMA locus is highly permissive to transcription initiation, and RNAPII occupancy in this region is simply a reflection of promiscuous unregulated RNAPII recruitment and initiation.

Our data demonstrating constitutive GTF, Mediator, and elongation factor occupancy throughout the first 6 kb of the PUMA locus merely indicates that this region is competent for transcription, and therefore does not discriminate among these hypotheses. However, our extensive RNA analysis demonstrates constitutive sense transcription of a large nonpolyadenylated, noncoding RNA encompassing the first half of the PUMA locus. The function, or lack thereof, of PUMA-TUF is, at present, unknown, and is the topic of ongoing research. In its simplest form, PUMA-TUF may represent nothing other than transcriptionally active promoters within PUMA giving rise to functionally elongating RNAPII complexes that pause/terminate within the first 200 bp of intron 3 just upstream of a 300-bp GC-rich sequence (>90% GC). Of note, GC-rich sequences have been demonstrated to pause polymerases under a variety of...
CTCF is described as a cornerstone of these novel chromatin architectures [Phillips and Corces 2009]. CTCF has repeatedly been documented to regulate the transcriptional state of specific gene loci by varied mechanisms, including but not restricted to (1) enhancer-blocking effects [Bell et al. 1999], (2) preventing spreading of heterochromatin into otherwise active genes [Witcher and Emerson 2009], and (3) regulation of promoter-proximal noncoding RNAs [Lefevre et al. 2008]. However, little is known about the function of intragenic CTCF occupancy found in up to ~30% of genes [Kim et al. 2007].

The new model of genome organization must also incorporate our ever-evolving understanding of chromatin architecture. The dated model of simple "open" euchromatin versus "closed" heterochromatin has given way to notions of interchromosomal interactions as well as long- and short-range chromatin loops, all occurring in the three-dimensional space of the nucleus [Lomvardas et al. 2006]. Recurrently, the insulator-binding protein CTCF is described as a cornerstone of these novel chromatin architectures [Phillips and Corces 2009]. CTCF has repeatedly been documented to regulate the transcriptional state of specific gene loci by varied mechanisms, including but not restricted to (1) enhancer-blocking effects [Bell et al. 1999], (2) preventing spreading of heterochromatin into otherwise active genes [Witcher and Emerson 2009], and (3) regulation of promoter-proximal noncoding RNAs [Lefevre et al. 2008]. However, little is known about the function of intragenic CTCF occupancy found in up to ~30% of genes [Kim et al. 2007].

Our studies indicate that the peculiar transcriptional organization of the PUMA locus goes hand in hand with an unusual chromatin landscape. Histone marks that are associated with transcriptionally active chromatin are present throughout the first half of the PUMA locus (i.e., H3K4me3 and H3K9Ac). Flanking this permissive chromatin region are marks associated with a repressed transcriptional state (i.e., H3K9me3), illustrating drastically different chromatin states within the body of the gene. Dual sites of CTCF and Cohesin complex occupancy mark this DNaseI-accessible domain within PUMA, and CTCF knockdown leads to a specific increase in basal PUMA expression. These results ascribe a novel genespecific function to intragenic CTCF and, by extension, Cohesin complex] occupancy by regulating gene expression through inhibition of transcriptional elongation. This establishes their important role within the p53 pathway, as it reduces the expression of a highly potent proapoptotic gene in the absence of cellular stress. We are left to ponder the function of the observed intragenic CTCF and Cohesin complex occupancy downstream from the transcriptional start sites on p21 and other p53 target genes. Interestingly, a recent study demonstrated that intragenic CTCF binding functions as a "late" checkpoint in early stages of elongation at the U2 snRNA and β-actin genes [Egloff et al. 2009]. This leaves open the possibility that CTCF and the Cohesin complex may function in a similar fashion at p21 and other p53 target genes. Additionally, it is possible that CTCF occupancy sites are looping to extragenic enhancers or regulatory sites, even perhaps in a cell type-specific manner, as demonstrated recently for the β-globin locus [Hou et al. 2010].

Taken together, the results presented here demonstrate a novel gene-specific regulatory mechanism for a critical gene involved in p53-dependent apoptosis, provide insight into a potential therapeutic strategy to modulate cell fate choice to p53 activation, and generate a model system to understand the unexpected complexity of eukaryotic genomes.

Materials and methods

ChIP assays

All ChIP analysis was performed as detailed in Gomes et al. [2006]. Briefly, HCT116 cells were grown to ~60% confluency, treated with 5-FU (375 μM, Sigma-Aldrich) for 8 h, and fixed with 1% formaldehyde, and then whole-cell lysates were prepared. Protein lysate (1 mg) was subjected to ChIP with the indicated antibodies [Supplemental Table S1], followed by DNA purification. ChIP-enriched DNA was analyzed with Q-PCR as
detailed in Gomes et al. (2006) with the indicated primer sets (Supplemental Table S2).

**Q-RT–PCR reactions**

All RT and Q-PCR analyses were performed as detailed in Gomes et al. (2006). Briefly, total RNA was prepared from HCT116 control and 5-FU-treated cells with Trizol (Invitrogen) following the manufacturer’s instructions. RT reactions were performed with SuperScript II (Invitrogen) using the random hexamer protocol following the manufacturer’s instructions. The resulting cDNAs were subjected to Q-RT–PCR with the indicated primer sets [Supplemental Table S3]. Values were normalized to those of 18S RNA. For Figure 4B (top), 50 ng of genomic DNA or 5 ng of cDNA was used in standard PCR reactions for 35 cycles using the indicated primers [Supplemental Table S3]. PCR products were resolved in 2% agarose gels. For Figure 4E, 50 ng of genomic DNA or 2 ng of cDNA was used in standard PCR reaction for 34 cycles using the indicated primers [Supplemental Table S3]. PCR products were resolved in 2% agarose gels.

**Protein immunoblot analysis**

Typically, 10 μg of total protein extract was loaded onto 12% SDS-PAGE and transferred to PVDF membranes. Blots were probed with the indicated primary antibodies [Supplemental Table S1], and were developed with peroxidase-conjugated secondary antibodies [Santa Cruz Biotechnologies] and ECL detection reagents [GE Healthcare].

**Flow cytometry analyses**

For apoptotic index assays, cells were harvested by trypsinization, washed with PBS, and resuspended in Annexin V Binding Buffer (10 mM HEPES at pH 7.4, 140 mM NaCl, 2.5 mM CaCl₂). Cells were then stained with Annexin V-fluorescein conjugate [Invitrogen] and propidium iodide, following the manufacturer’s instructions. Cells were analyzed on a Beckton-Dickinson FACScan instrument.

**Strand-specific exon arrays**

Total RNA was prepared from control and 5-FU-treated (12 h) HCT116 cells with Trizol reagent (Invitrogen) following the manufacturer’s instructions. Each RNA sample was subjected to rRNA removal using the RiboMinus Human/Mouse Transcriptome Isolation Kit [Invitrogen], and sense strand cDNA was synthesized using the GeneChip WT [Whole Transcript] Sense Target Labeling kit [Invitrogen] according to the manufacturer’s instructions. Sense strand cDNA was then fragmented and biotin-labeled using the GeneChip WT Terminal Labeling kit [Invitrogen] following the manufacturer’s instructions. Samples were hybridized to Affymetrix GeneChip Human Exon 1.0 ST arrays. Each array was washed and stained using a GeneChip Hybridization, Wash, and Stain kit [Invitrogen]. Arrays were scanned with an Affymetrix scanner 3000 7G. Exon array data were analyzed with Partek Genomics Suite software [Partek, Inc.] using the built-in robust multiarray analysis (RMA) algorithm.

**Strand-specific Northern blots**

Total RNA was prepared from control and 5-FU-treated (16 h) HCT116 cells with Trizol reagent (Invitrogen) following the manufacturer’s instructions and was suspended in RNA loading buffer (48% formamide, 6% formaldehyde, 50 mM HEPES, 1 mM EDTA, 8% glycerol), and 10 μg of each RNA sample was resolved in denaturing 1.5% agarose gels (6% formaldehyde, 50 mM HEPES, 1 mM EDTA). RNA was transferred overnight to Hybond-N* membrane [GE Healthcare] with capillary transfer and then cross-linked with ultraviolet light. Sense strand-specific 80-mer oligonucleotide probes (Supplemental Table S4) were end-labeled using T4-PNK [New England Biolabs] and γ-32P-ATP. Membranes and probes were hybridized overnight in Church buffer (1 mM EDTA, 0.5 M phosphate buffer at pH 7.2, 7% SDS) at 65°C. Membranes were washed twice for 20 min with 1× SSC/0.1% SDS and twice for 10 min with 0.5× SSC/0.1% SDS. Membranes were exposed to storage phosphor screens [Invitrogen] and were scanned on a Storm 860 [Invitrogen]. For Figure 4D, polyA⁺ RNA was prepared with Oligotex resins [Qiagen] following the manufacturer’s instructions. One microgram of total or polyA⁺ RNA was subjected to Northern blot analysis as described above.

**shRNA-mediated knockdown**

shRNA-mediated knockdowns were performed essentially as in Knuesel et al. (2009). Briefly, in HEK293FT cells, pL3.7 control and CTCF shRNA (5′-GGACAGTGTTGACAAT3A′) lentiviruses were produced with generation III packaging vectors, while pGIPZ control and CTCF shRNA (5′-GAGTCAATTGAGAA CATT-3′) plasmids were produced with the Trans-Lentiviral Packaging System [Open Biosystems]. HCT116 cells were transduced with virus for 8 h in the presence of polybrene (8 μg/mL, Sigma-Aldrich). Cultures were then selected for 3–4 d with puromycin (10 μg/mL, Sigma-Aldrich) and then harvested for the experiments detailed in Figures 6 and 7.

**Acknowledgments**

We are thankful to members of the Espinosa laboratory for contributing ideas and reagents. This work was supported by NIH grant RO1-CA117907. J.M.E. is a Howard Hughes Medical Institute Early Career Scientist. N.P.G. was supported in part by NIH training grant T32GM07135.

**References**


Stewart MD, Li J, Wong J. 2005. Relationship between histone H3 lysine 9 methylation, transcription repression, and


