Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies

Lisa L. Hall
University of Massachusetts Medical School

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/oapubs

Part of the Biochemistry, Biophysics, and Structural Biology Commons, Cell and Developmental Biology Commons, and the Genetics and Genomics Commons

Repository Citation

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License. This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Open Access Publications by UMass Chan Authors by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies

Graphical Abstract

Highlights
- HSATII misregulation is common in cancers and is locus and tumor specific
- Demethylation of HSATII DNA at the 1q12 mega-satellite sequesters PRC1 into CAP bodies
- Large RNA foci from other specific HSATII loci accumulate MeCP2 into CAST bodies
- Sequestration of master regulators by HSATII may further compromise the epigenome

Authors
Lisa L. Hall, Meg Byron, Dawn M. Carone, ..., Andrew Fischer, Peter Jones, Jeanne B. Lawrence

In Brief
Satellite II is a prominent but poorly studied feature of human genomes. Hall et al. show that HSATII DNA and RNA can sequester PRC1 and MeCP2. In cancer, PRC1 bodies form on the demethylated 1q12 mega-satellite, while MeCP2 bodies form on HSATII RNA, potentially leading to further changes in the epigenome.

Accession Numbers
GSE94973
GSM2492610
GSM2492611
GSM2492612
GSM2492613
Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies

Lisa L. Hall,1,4 Meg Byron,1,4 Dawn M. Carone,1,6 Troy W. Whitfield,1,3 Gayle P. Pouliot,1,5 Andrew Fischer,2 Peter Jones,1,7 and Jeanne B. Lawrence1,8,*

1Department of Cell and Developmental Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA
2Department of Pathology, University of Massachusetts Medical School, Worcester, MA 01655, USA
3Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA
4Co-first author
5Present address: Children’s Hospital, Harvard Medical School, Boston, MA 01225, USA
6Present address: Department of Biology, Swarthmore College, Swarthmore, PA 19081, USA
7Present address: Department of Pharmacology, University of Nevada School of Medicine, Reno, NV 89557, USA
8Lead Contact
*Correspondence: jeanne.lawrence@umassmed.edu
http://dx.doi.org/10.1016/j.celrep.2017.02.072

SUMMARY

This study reveals that high-copy satellite II (HSATII) sequences in the human genome can bind and impact distribution of chromatin regulatory proteins and that this goes awry in cancer. In many cancers, master regulatory proteins form two types of cancer-specific nuclear bodies, caused by locus-specific deregulation of HSATII. DNA demethylation at the 1q12 mega-satellite, common in cancer, causes PRC1 aggregation into prominent Cancer-Associated Polycomb (CAP) bodies. These loci remain silent, whereas HSATII loci with reduced PRC1 become derepressed, reflecting imbalanced distribution of UbH2A on these and other PcG-regulated loci. Large nuclear foci of HSATII RNA form and sequester copious MeCP2 into Cancer-Associated Satellite Transcript (CAST) bodies. Hence, HSATII DNA and RNA have an exceptional capacity to act as molecular sponges and sequester chromatin regulatory proteins into abnormal nuclear bodies in cancer. The compartmentalization of regulatory proteins within nuclear structure, triggered by demethylation of “junk” repeats, raises the possibility that this contributes to further compromise of the epigenome and neoplastic progression.

INTRODUCTION

Epigenetic changes are recognized as a major player in tumorigenesis (Feinberg, 2014; Sandoval and Esteller, 2012). Although attention has been focused on methylation and silencing of tumor suppressor genes, this often occurs concomitant with loss of methylation from other genomic regions, including pericentric satellites. High-copy tandem satellites repeat families (alpha, beta, SATI, II, III) constitute ~15% of the human genome. Alpha-satellite (α-SAT) is at the centromere of all chromosomes, and satellite III (HSATIII) on chromosome 9 (chr9) is linked to the heat shock response and nuclear stress bodies (Biamonti and Vourc’h, 2010). However, many repeats are considered mere evolutionary relics and have been poorly studied. Perhaps the best example of this is human satellite II (HSATII), an exceptionally high-copy repeat with no raison d’etre. Small blocks of HSATII are found on the pericentromeres of 11 human chromosomes (chr2, 5, 7, 10, 13, 14, 15, 17, 21, 22, and Y), but 1q12 on chr1, and to a lesser extent 6q11 on chr16, carries very large (~5–6 Mb) blocks of this ~26-bp tandem repeat. There is frequent loss of methylation of HSATII in cancer (Ehrlich, 2009), but any significance of this is unknown. HSATII remains one of the most prominent but poorly studied features of the human genome.

Pericentric heterochromatin is regulated by the polycomb group (PcG) complexes PRC1 and PRC2, which also control most early developmental pathways and are commonly misregulated in cancer (Simon and Kingston, 2009). Pericentric heterochromatin is generally silent but can be expressed in relation to cell cycle, heat shock, and stress (Hall et al., 2012). When we found an unexpected association of the BRCA1 tumor suppressor with satellite DNA during its replication (Pageau and Lawrence, 2006), we hypothesized that satellites might be expressed in cancer (Pageau et al., 2007). Studies subsequently reported pericentric satellite expression in several cancers (Eymery et al., 2009; Ting et al., 2011; Zhu et al., 2011), but many questions remain as to the extent, specificity, and significance of this. Does it reflect broad non-specific heterochromatic instability across all satellites, or might it show family or locus specificity? How does it arise, and, perhaps most importantly, does it have any impact on the cell? Study of satellites is limited using standard genomic approaches, as evidenced by the lack of the extraordinarily large 1q12 HSATII in the human genome database. Here, we use an approach to study satellite DNA and RNA, which preserves molecular information within cell
and nuclear/chromosome structure and allows analysis not only of specific repeat families, but also of specific loci of a given family within single cells.

Our results support a paradigm for the significance of changes to satellite heterochromatin: the million or more copies of HSATII per cell have a pronounced capacity, at both the DNA and RNA levels, to absorb and impact the distribution of key epigenetic regulators across the nuclear genome. This goes awry in cancer and is linked to the common demethylation of HSATII, and the “mega-satellite” at 1q12 may play a singular role.

RESULTS

Aberrant Nuclear Foci of Satellite RNA Are Common in a Wide Array of Tumor Types

We previously showed that a CoT-1 (highly repetitive) DNA probe (Britten and Kohne, 1968) detects an evenly distributed RNA signal across the euchromatic nucleoplasms by fluorescence in situ hybridization (FISH) (Figure 1A) (Hall et al., 2002, 2014). However, we noted some cell samples also contained very prominent CoT-1 RNA foci (Figures 1B, 1C, S1A, and S1B), which can be quite large (compare to XIST RNA; Figure 1D). In 22 cell lines surveyed, CoT-1 RNA foci were in 11 of 14 neoplastic lines but none of eight normals (Table S1).

As CoT-1 is heterogeneous, probes to different repeat families (see Supplemental Information) were used to identify specific repeats within the RNA foci. Although interspersed LINE1 or Alu (SINE) repeats were not present in CoT-1 RNA foci (Figures S1C and S1D), z-SAT (Figures S1E and S1F) and HSATII were (Figure 1E). These RNA foci are nuclear restricted, and several observations affirm these foci are single-stranded RNA: detection without DNA denaturation (which will not detect even these highly abundant DNA sequences) (Lawrence et al., 1988), loss of signal with RNase (Figures S1G, S1H, S2D, and S2E) or NaOH, specificity for certain cell lines, absence from mitotic chromosomes, and release of these HSATII RNA “bodies” into the cytoplasm of mitotic cells (Figures S6G and S2F).

We examined z-Sat and HSATII expression by RNA FISH in 14 cancer cell lines and eight non-cancer lines, nine human effusion samples, and 44 human solid tumors (with matched normal where available) for a wide array of tumor types (Table S1). Given that RNA preservation in pathology samples can be variable, we used poly(A) RNA as a positive control. Satellite RNA foci were detectable in over half of the tissue samples (solid tumor and effusion) (Figures 1F–1H) and ten of 14 cancer lines but in none of the normal tissue or lines (Table S1). Thus, aberrant nuclear satellite RNA foci are common in human cancers in vivo and in vitro.

**Satellite Dysregulation Is Family Specific, Primarily HSATII, and Not Tied to BRCA1 Status**

Does this misregulation reflect a generalized failure to maintain silencing on all satellites or preferentially impact a specific repeat family or locus? This has implications for whether misregulated satellites are a consequence or contributor to neoplasia.

The conserved z-SAT repeat is easily distinguished on all chromosomes. However, due to sequence similarity we tested a number of probes and labeling strategies to identify probes (and conditions) that discriminate HSATIII on chr9 from HSATII on other chromosomes (Figures S1I–S1M). The HSATII probe, and an antibody to Heat shock Factor 1 (HSF1), confirmed that the RNA expressed in these unstressed cancer cells was HSATII (Figures S1I–S1O) (details in Supplemental Information).

Numerous small dim nuclear foci of human z-SAT RNA were visible in all normal cell lines examined, including proliferative fibroblasts and post-mitotic myotube nuclei (Figures S2A–S2F; Table S1). Since our initial observations, centric satellite RNAs were reported at very low levels in normal cells (Bierhoff et al., 2014), supporting the accuracy and sensitivity of our in situ approach. However, these small foci were still distinct from the larger z-Sat foci that overlapped CoT-1 RNA foci in some cancer cells (Figures S2G–S2J).

In contrast to z-sat, HSATII RNA is essentially undetectable in normal cells using multiple HSATII probes (see Supplemental Information) (Figures S1I–S1N; Table S1). Normal cells only occasionally had a tiny pinpoint of HSATII RNA fluorescence detectable with digital imaging. Conversely, HSATII RNA foci were quite large in cancer nuclei and distinct enough to identify a single cell as cancerous. This was affirmed by quantitative digital imaging (Figures 1L–1N) and showed U2OS cells with ~175-fold more HSATII RNA than normal cells (Figure 1N). The intensity and size (~0.4–1 μm) of these RNA conglomerations (Figure 1D; Table S3) indicate extremely high copies of the 26-bp HSATII sequence.

Examining the frequency of z-Sat and HSATII overexpression, we found z-SAT RNA foci were seen in only two of nine cancer lines (HT1080 and MDA-MB-436) (Figures S1E and S1F; Table S1) and multiple probes (Supplemental Information) were used to identify specific repeats within the RNA foci. Although interspersed LINE1 or Alu (SINE) repeats were not present in CoT-1 RNA foci, we noted some cell samples also contained very prominent CoT-1 RNA foci (Figures 1B, 1C, S1A, and S1B), which can be quite large (compare to XIST RNA; Figure 1D). In 22 cell lines surveyed, CoT-1 RNA foci were in 11 of 14 neoplastic lines but none of eight normals (Table S1).

As CoT-1 is heterogeneous, probes to different repeat families (see Supplemental Information) were used to identify specific repeats within the RNA foci. Although interspersed LINE1 or Alu (SINE) repeats were not present in CoT-1 RNA foci (Figures S1C and S1D), z-SAT (Figures S1E and S1F) and HSATII were (Figure 1E). These RNA foci are nuclear restricted, and several observations affirm these foci are single-stranded RNA: detection without DNA denaturation (which will not detect even these highly abundant DNA sequences) (Lawrence et al., 1988), loss of signal with RNase (Figures S1G, S1H, S2D, and S2E) or NaOH, specificity for certain cell lines, absence from mitotic chromosomes, and release of these HSATII RNA “bodies” into the cytoplasm of mitotic cells (Figures S6G and S2F).

We examined z-Sat and HSATII expression by RNA FISH in 14 cancer cell lines and eight non-cancer lines, nine human effusion samples, and 44 human solid tumors (with matched normal where available) for a wide array of tumor types (Table S1). Given that RNA preservation in pathology samples can be variable, we used poly(A) RNA as a positive control. Satellite RNA foci were detectable in over half of the tissue samples (solid tumor and effusion) (Figures 1F–1H) and ten of 14 cancer lines but in none of the normal tissue or lines (Table S1). Thus, aberrant nuclear satellite RNA foci are common in human cancers in vivo and in vitro.

**Satellite Dysregulation Is Family Specific, Primarily HSATII, and Not Tied to BRCA1 Status**

Does this misregulation reflect a generalized failure to maintain silencing on all satellites or preferentially impact a specific repeat family or locus? This has implications for whether misregulated satellites are a consequence or contributor to neoplasia.

The conserved z-SAT repeat is easily distinguished on all chromosomes. However, due to sequence similarity we tested a number of probes and labeling strategies to identify probes (and conditions) that discriminate HSATIII on chr9 from HSATII on other chromosomes (Figures S1I–S1M). The HSATII probe, and an antibody to Heat shock Factor 1 (HSF1), confirmed that the RNA expressed in these unstressed cancer cells was HSATII (Figures S1I–S1O) (details in Supplemental Information).

Numerous small dim nuclear foci of human z-SAT RNA were visible in all normal cell lines examined, including proliferative fibroblasts and post-mitotic myotube nuclei (Figures S2A–S2F; Table S1). Since our initial observations, centric satellite RNAs were reported at very low levels in normal cells (Bierhoff et al., 2014), supporting the accuracy and sensitivity of our in situ approach. However, these small foci were still distinct from the larger z-Sat foci that overlapped CoT-1 RNA foci in some cancer cells (Figures S2G–S2J).

In contrast to z-sat, HSATII RNA is essentially undetectable in normal cells using multiple HSATII probes (see Supplemental Information) (Figures S1I–S1N; Table S1). Normal cells only occasionally had a tiny pinpoint of HSATII RNA fluorescence detectable with digital imaging. Conversely, HSATII RNA foci were quite large in cancer nuclei and distinct enough to identify a single cell as cancerous. This was affirmed by quantitative digital imaging (Figures 1L–1N) and showed U2OS cells with ~175-fold more HSATII RNA than normal cells (Figure 1N). The intensity and size (~0.4–1 μm) of these RNA conglomerations (Figure 1D; Table S3) indicate extremely high copies of the 26-bp HSATII sequence.

Examining the frequency of z-Sat and HSATII overexpression, we found z-SAT RNA foci were seen in only two of nine cancer lines (HT1080 and MDA-MB-436) (Figures S1E and S1F; Table S1).
Table S1) and 25% of tumor samples, whereas HSATII RNA was found, in eight of nine cancer lines (70%–100% of cells) and about half of tumor and effusion samples (Table S1). Moreover, 50% of tumors with satellite RNA foci had HSATII exclusively (11/22 samples tested for both), and only 9% had z-SAT alone (and 41% with both). Zhu et al. (2011) reported z-SAT RNA expression in BRCA1+/− breast cancers and concluded loss of BRCA1 specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition). However, BRCA1+ tumors were never examined, and our results show specifically causes this (due to loss of UbH2A deposition).

**PRC1 Redistributes into Large Cancer-Associated Polycomb Bodies on 1q12, which Remains Repressed**

We examined Polycomb Group (PcG) proteins, known to repress heterochromatin (including satellites), and that are implicated in cancer pathogenesis (Sauvageau and Sauvageau, 2010). Staining for BMI-1, RING1B and Phc1 (PRC1), and EED and EZH2 (PRC2) revealed PRC1 is aberrantly distributed in cancer cells (Figures 2 and S2K) but generally not PRC2 (Figure S2L). BMI-1 labels a few prominent nuclear bodies in 88% of neoplastic lines and 44% of primary tumor samples but not any normal tissues/lines (Figures 2A–2G; Table S2). Since components of PRC1 can vary with cell type, these numbers (using BMI-1) may underestimate the frequency of PRC1 in these bodies. The large (0.4–1.5 μm) BMI-1 “bodies” were often 20-fold brighter than the surrounding nucleoplasm, whereas most normal cells had hundreds of tiny punctates (Figures 2H–2L). The most prominent BMI-1 foci in non-cancer cells (e.g., immortalized retinal pigment epithelium [RPE] cells) had at most a “contrast ratio” of just 4:1 (Figures 2I and 2K). Importantly, the large cancer-associated PRC1 bodies consistently (~100%) and precisely co-localize with 1q12 HSATII DNA (and to a lesser extent HSATII on chr16q11) (Figure 3A).

Our findings establish that these large PRC1 aggregates on 1q12 are not normal nuclear structures but an epigenetic hallmark of cancer. Numerous small PcG puncta, widely dispersed in the nucleoplasm (termed “PcG bodies”), are found in all normal cells from *Drosophila* to humans and are reported to help regulate gene expression (Bantignies and Cavalli, 2011; Geyer et al., 2011). PcG bodies have often been studied in cancer cells (widely assumed to model normal nuclear structures) (Dundr, 2012), where it was noted cancer and normal cells might differ in the number and size of their “PcG bodies” (Saurin et al., 1996). However, our analysis establishes that the few prominent bodies in cancer are a completely different structure, with different biological significance. To avoid confusion with the small, numerous, and widely distributed PcG puncta in normal human cells, often referred to as “PcG bodies,” we refer to these large conglomerations of PRC1 proteins that specifically “cap” the HSATII locus at 1q12 in cancer cells as “CAP” bodies, for “cancer-associated Polycomb” bodies.

We next examined the relationship between HSATII RNA foci and PRC1 accumulation into bodies, with two alternate possibilities: HSATII RNA might recruit PRC1 to 1q12, as do other ncRNAs, such as XIST, or enrichment of PRC1 at 1q12 might maintain repression, while other loci, in the nucleoplasm with lower levels of PRC1, may become derepressed. Results show that the latter consistently occurs; HSATII RNA foci and CAP bodies are spatially separated (0% overlapping, 6% adjacent, and 94% no association) (Figures 3B–3D), and sequential hybridization to HSATII RNA and DNA (Byron et al., 2013) indicates that RNA foci emanate from small or medium HSATII DNA loci but not the largest HSATII loci of 1q12 (Figure 3E). Importantly, when CAP bodies and HSATII RNA foci are both detected in a given tumor, they are typically in the same nuclei, suggesting a relationship. For example, in breast tumor #2334T, 80% CAP positive nuclei also contained RNA foci (e.g., Figure 3D), which is likely an underestimate since nuclei in the tissue are sliced.

**Locus-Specific HSATII Sequence Subtypes Exist and May Underlie Differences in PRC1 Accumulation or Deregression in Cancer**

The different biological behaviors of HSATII loci raise the question of whether there are variations in HSATII sequences at different loci. Since most HSATII sequence is not in the human genome build (hg19) (particularly 1q12), and it is difficult to establish locus specificity of small tandem repeats by sequence data, in situ hybridization allows us to investigate whether specific sequences are enriched at specific loci.

As detailed in the Supplemental Information, we examined a series of probes representing different permutations of divergent HSATII sequences, under varying hybridization stringencies. Because the 1q12 locus is so large, most sequences found on other chromosomes cross-hybridize to 1q12 and vice versa. However, from the larger pUC1.77 cloned sequence from 1q12 we identified a smaller 160-bp PCR-generated sequence (Sat2-160) that showed specific hybridization to 1q12, with slight detection of 16q11. This probe did not detect HSATII RNA and correlated strongly with BMI-1 aggregates (Figures S2M–S2O), consistent with the above results. While there is no known consensus sequence for mammalian PRC1 binding, this demonstrates that distinct subsets of HSATII sequences characterize different loci, which likely contributes to distinct biological behavior.

A separate question was whether we could identify specific HSATII sequences/loci that are expressed. As summarized in the Supplemental Information, we identified an expressed sequence by RT-PCR (Figure S3A) that only labeled one to two RNA foci in U2OS nuclei (by RNA FISH), which have an unusually high number of RNA foci detected with the broader Sat2-24 LNA probe (Figures 3E and 3F; Table S3), suggesting locus specificity of the PCR probe. This probe also labeled the most typical one to two RNA foci detected by the permissive LNA probe in tumor
tissues (Figure 3G). We tentatively mapped the sequence to chromosome 7 (using hg19); however, since the probe labeled more than one chromosome by DNA FISH, specific centromere probes were used to confirm expression from chromosome 7 (Figure 3H). Of five tumor samples that scored high for HSATII RNA, three expressed this sequence (now called Sat2-7) (three positive breast tumors: #0853, #2004, and #2334, and two negative stomach tumors: #2539 and #2210). We also found HSATII sequences from chr7 are also one of the more commonly expressed sequences in a pancreatic cancer deep sequencing database (SRX056914) (Figure S3B), suggesting the chr7 locus may be commonly misregulated in some cancers.

Figure 2. BMI-1 Aggregates in Large Aberrant Foci that Form Nuclear Bodies
(A and B) BMI-1 protein accumulates into bodies in cancer cells.
(C and D) Normal fibroblasts only have faint punctate nucleoplasmic signal. Red channel enhanced at right. Exposures are the same for (A)–(C).
(E) Cancer cell lines contain BMI-1 bodies, while non-cancer lines do not.
(F) Normal human tissue with uniform punctate distribution of BMI-1.
(G) The matched tumor sample contains BMI-1 bodies.
(H–J) Low nucleoplasmic BMI-1 in fibroblasts (H), slightly higher levels in telomerase immortalized RPE cells (I), and large accumulations in cancer cells (J), with concomitant reduction in nucleoplasmic signal. (H–J are the same scale.)
(K and L) BMI-1 line scans of the same cells in (H)–(J).
These findings demonstrate that there is a locus-specific distribution of HSATII sequence subtypes on different loci that may underlie differences in biological behavior. Further, the dysregulation of HSATII in cancer samples is not stochastic nor indiscriminate but often shows locus and sequence specificity for a given tumor (e.g., Table S3). The HSATII locus at 1q12 consistently nucleates large CAP bodies and is marked by specific subtypes of HSATII not detected in RNA, whereas the Sat2-7 sequence we identified on chr7p11 appears to be specifically expressed in several tumors studied here.

**HSATII Changes Link to Broader Epigenetic Instability in Cancer from DNA Demethylation to UbH2A Distribution**

We investigated whether changes in HSATII and CAP body formation relate to broader epigenetic instability in cancer and found evidence they do, both in terms of their cause but also likely their consequences. DNA methylation is common in cancer, and HSATII demethylation at 1q12 is a frequent and often early change (Ehrlich, 2009). We therefore tested the effects of inducing DNA demethylation on HSATII expression or potentially redistribution of PRC1.

We first examined cells of the rare and fatal immunodeficiency-centromere instability-facial anomalies (ICF) syndrome, where a DNMT3B mutation causes HSATII demethylation (Walton et al., 2014), and found PRC1 accumulated in large bodies that “cap” 1q12 in 61% of the ICF fibroblasts (Figures 4E, S3C, and S3D). We next treated normal human fibroblasts with 5-aza-2'-deoxycytidine, which we previously confirmed demethylates HSATII in these cells (Swanson et al., 2013). Remarkably, within 24 hr PRC1 accumulates into one to two large bodies over 1q12 (Figures 4A–4C). Longer treatment increased the number of cells affected (day 1 = 15%, day 3 = 52%, day 8 = 80%), and the number of bodies/cell (from two to four, representing accumulation on chr1q12). After several days, a subset of cells (5%–10%) began to express HSATII RNA from other loci (Figure 4D). Thus, loss of cytosine methylation on HSATII triggers massive recruitment of repressive PRC1 to 1q12, while other loci eventually become derepressed.

Transient activity of PRC2 HMTase EZH2 is required to recruit PRC1 (Hernández-Muñoz et al., 2005). No consistent accumulation of EZH2 was seen on 1q12 in cancer cells (Figure S2L) nor following 5-aza2d treatment (Figure 4F), but an increase in nucleoplasmic H3K27me3 was seen (not limited to nor enriched on pericentromeres) (Figures 4G and 4H), suggesting PRC2 activation. Inhibition of EZH2 (3-deazaneplanocin) during 5-aza-2d treatment showed reduced H3K27me3 staining (confirming EZH2 inhibition) (Figure 4I) and slowed/reduced CAP body

**Figure 3. PRC1 Aggregates on HSATII DNA at 1q12 Maintaining Repression, while Other Loci Become Derepressed**

(A) The pUC1.77 probe labels HSATII on chr1 and some on chr16 (arrows) and correlates with BMI-1 “CAP” bodies. Channels from outlined region are separated at right.

(B and C) CAP bodies and HSATII RNA are mutually exclusive. Red and green channels of (B) at right.

(D) CAP bodies and HSATII RNA foci in breast tumor tissue.

(E) HSATII RNA is expressed from small HSATII loci and not from the large 1q12 loci. Small loci are visible when DNA signal (of outlined region) is separated and enhanced at right. HSATII RNA accumulates adjacent to DNA loci (arrows).

(F and G) Sat2-7 probe labels only one to two RNA foci in U2OS cells (F) or in breast tumor #2334T (G).

(H) chr7-specific centromere probe verifies Sat2-7 RNA is expressed from chr7.
formation (Figure 4J), suggesting EZH2/PRC2 activity is likely involved in CAP body formation.

Corralling PRC1 into CAP bodies likely reduces its level elsewhere, which can be seen by immunofluorescence (IF) in U2OS cells (Figure 2J), at low (10×) magnification in tumor sections (Figures 5A and 5B) and can be measured by line scan, when comparing neighboring nuclei with and without CAP bodies (Figures 5C–5E; renal tumor). These findings point to a biological significance for satellite demethylation in human cancers, due to the marked capacity of this extremely high-copy repeat at 1q12, to act as a molecular “sponge,” thereby impacting compartmentalization and accessibility of important nuclear factors.

To assess potential downstream effects of PRC1 redistribution into CAP bodies, we examined UbH2A distribution (laid down by PRC1), by chromatin immunoprecipitation (ChIP), in cancer and normal cells. Reads mapping to HSATII increased sharply in cancer cells, from 8% in normal Tig-1 fibroblasts to 37% in U2OS osteosarcoma cells (Figure 5F). This is consistent with PRC1 accumulation at 1q12, which would dominate the HSATII sequence population. The distribution of UbH2A on 1q12 (Sat2-160) versus chr7 (Sat2-7) satellites in normal cells is similar (since both loci are repressed), whereas in U2OS cancer cells the 1q12 sequences become disproportionately enriched for UbH2A over the expressed chr7 sequence (Figures 5G and S3E). Throughout the broader epigenome, UbH2A distributed more uniformly in Tigs (Figure 5H), whereas in U2OS more sequences deviate from average, particularly at the low ends (Figure 5I). Gene ontology analysis for genes with reduced UbH2A (4-fold or

Figure 4. CAP Bodies Are Induced by DNA Demethylation and PRC2 Activity
(A and B) Fibroblasts treated with 5aza-d form CAP bodies. Red channel is separated in (B).
(C) CAP bodies form over 1q12 in demethylated fibroblasts. Insets show colors separated for top body.
(D) HSATII is expressed following longer 5aza-d treatment. Inset shows green channel of outlined region.
(E) CAP bodies are present in ICF cell nuclei. Insets show color channels separated for CAP body.
(F) EZH2 (PRC2) does not aggregate in CAP bodies following 5-aza.
(G) Low-level H3K27me in controls, with enrichment on the inactive X chromosome.
(H) Increased H3K27me following 5-aza treatment.
(I) H3K27me levels reduced with simultaneous treatment with EZH2-inhibitor and 5-aza. The exposures are equal for (G)–(I).
(J) Number of cells with CAP bodies on days 3 and 5 following treatment with 5-aza (D3-A and D5-A) or 5-aza and EZH2-inhibitor (D3-A/Ei and D5-A/Ei).
Figure 5. CAP Bodies Sequester PRC1 Affecting Genomic Distribution of UbH2A

(A) Low-magnification (10x) view of BMI-1 in tumor tissue.
(B) Magnification of area in box of (A) shows cells with CAP bodies (arrows) and reduced nucleoplasmic BMI-1.
(C–E) Tumor cells with CAP bodies have reduced nucleoplasmic BMI-1 compared to neighboring cells lacking CAP bodies. Green channel separated in (D) and line scan of cell 1 and 2 (E) are shown.
(F) ChIP-seq: total HSATII reads with UbH2A in cancer versus normal cells.
(G) ChIP-seq: UbH2A levels on expressed (Sat2-7) versus repressed (1q12) HSATII sequences in normal and cancer cells (total reads Figure S3E).
(H and I) ChIP-seq: distribution of UbH2A across the genome in normal (H) and cancer (I) cells.
(J) ChIP-PCR: UbH2A enrichment on 1q12, HoxC5, and β-actin loci before and after 5aza treatment.
(K) RING1B/RNF2 on the EGLN3 gene (involved in hereditary breast and ovarian cancer) in cancer cells and normal foreskin and breast fibroblasts.
greater) in cancer cells suggested effects on genes involved in neural differentiation (Table S4A), consistent with the known role of BMI-1 (Corley and Kroll, 2013). Ting et al. also noted increased neural gene expression, a known poor prognostic indicator, was increased in cancers expressing HSATII (Ting et al., 2011). We also examined PRC1 distribution in published ChIP sequencing (ChIP-seq) datasets for RING1B/RNF2 (Pemberton et al., 2014; Rai et al., 2015) and found an increase in HSATII reads pulled down in cancer compared to normal and proportionally more on 1q12 (Sat2-160) than chr7 (Sat2-7) (Figure S3F). The genes that lost RING1B in cancer were also enriched for genes important to carcinogenesis (Table S4B; Figure 5K), illustrating how redistribution of PRC1 away from canonical genes could contribute to tumorigenesis (Wang et al., 2015).

We also used ChIP-PCR to examine UbH2A levels on three specific sequences in fibroblasts induced to form CAP bodies with 5-aza. UbH2A was again clearly increased on the 1q12 (Sat2-160), whereas HoxC5, an established PcG-regulated gene, demonstrated a corresponding sharp loss of UbH2A, with no significant change in the β-actin housekeeping gene (Figure 5J). These results further support that redistribution of PRC1 proteins into CAP bodies is associated with reduced access to repressive factors elsewhere, impacting UbH2A levels and making genes more vulnerable to derepression over time (e.g., HSATII at chr7).

**MeCP2 Accumulates with HSATII RNA Transcripts into Cancer-Associated Bodies**

Since HSATII DNA repeats can “sponge-up” PRC1, it follows that the similarly large HSATII RNA foci, containing likely millions of copies of this small repeat, might likewise sequester specific RNA-binding or regulatory factors, potentially similar to “Toxic repeat RNAs” in other human diseases (see Discussion).

Therefore, we examined whether HSATII RNA foci might accumulate MBNL1 (sequestered in repeat RNA in myotonic dystrophy type I), which it did not, nor did HSATII RNA overlap the PNC domain that sequesters proteins in many cancers (Wen et al., 2013). Surprisingly, we discovered the methyl-DNA binding protein, MeCP2, is found in large nuclear bodies that precisely overlap the HSATII RNA foci. These large MeCP2 foci (Figure 6A) are distinct from the broadly distributed small puncta in normal cells (Figure 6B) and can be seen homogeneously distributed throughout many tumors at low magnification (Figures S3G–S3I). We initially examined MeCP2 due to its capacity to bind methylated DNA, and, although MeCP2 foci were similar in size and number to both HSATII RNA foci and CAP bodies, detailed analysis clearly showed they were associated with the RNA, not DNA (Figures 6C–6F) (using two different antibodies and several cancer samples). For example, MeCP2 exactly overlaps HSATII RNA foci (Figure 6D), and since HSATII RNA foci are immediately adjacent to the DNA locus (Figure 3E) MeCP2 is not over the DNA. In addition, MeCP2 continues to overlap HSATII RNA foci when the “body” enters the cytoplasm at mitosis (Figures 6G and 6H).

We further show that MeCP2 and HSATII RNA co-immunoprecipitate from cancer cells by RNA immunoprecipitation (RIP). HSATII RNA was consistently enriched in MeCP2 pull-downs over immunoglobulin G (IgG) controls (Figure 6I), whereas α-SAT RNA was not. RIP with MBNL1 did not pull down HSATII RNA (confirming in situ results), while Sin3A, a transcriptional regulator that can complex with MeCP2 (Jones et al., 1998), did pull it down in some experiments. Using the RNA pulled down by MeCP2 (or SIN3A) as probes, we detected HSATII RNA foci in cancer cells by RNA FISH (Figures 6J and 6K). Thus, the RNA foci of HSATII are not inert but also can bind and impact distribution of epigenetic factors. We term these large round aggregates of HSATII RNA with MeCP2 “cancer-associated satellite transcript” (CAST) bodies and identify them as another unique hallmark of cancer cells (in vitro and in vivo).

MeCP2 is known to repress expression of LINE1 elements (Erwin et al., 2014), which are inappropriately expressed in cancer (Rodic and Burns, 2013). We recently showed that L1 sequences are a major part of the repeat-rich “CoT-1 RNA” that associates with euchromatin and may serve to maintain open chromatin (Hall et al., 2014). The peripheral heterochromatin-in-compartment in most normal cells characteristically lacks this L1-rich CoT-1 RNA (Figures S3J and S3K), reflecting a stable heterochromatic state. However, in cancer cells, CoT-1/L1 RNA is detected throughout this peripheral compartment, indicating compromise of this heterochromatin (Figures S3L and S3M; Pageau et al., 2007). In addition, we find U2OS tumor cells have more L1 RNA foci in nuclei compared to normal cells (Tig-1) (Figure 6L), consistent with other reports of L1 activation in cancers. Hence, the aberrant distribution of MeCP2 in cancer nucleus structure may be relevant to broader changes in cancer expression patterns.

Finally, we did not find evidence that HSATII RNA expression leads to massive tandem expansion of HSATII DNA loci, as a recent report suggests (Bersani et al., 2015). These authors postulate that HSATII cDNA (via reverse-transcription of RNA) is re-incorporated into and amplifies existing HSATII loci (e.g., up to 25-fold), since they saw no change in the number loci by DNA FISH. However, the relative size and intensity of DNA FISH signals is proportional to target size and much smaller differences (2- to 3-fold) are easily discernible (Lawrence et al., 1988). Using FISH and microfluorimetry to quantify the HSATII DNA signals in normal and cancer cells, we found no substantial differences in size, intensity, number of signals, or total HSATII signal (Figure 6M). This was quantified in U2OS osteosarcoma, but similar results were observed in other cancers. We detect a slight increase in total HSATII RNA in U2OS, due to an extra copy of the large 1q12 satellite (not tandem amplification), as evident in Figure 3A.

The model and legend in Figure 7 summarize and integrate several major findings in cancer nuclear structure and the epigenome, which we show are connected to complex and locus-specific biology of HSATII.

**DISCUSSION**

This work demonstrates multiple properties of cancer cell nuclear structure, with implications for epigenetic changes in cancer and at the same time points to the unanticipated biological importance of HSATII DNA, a large, unexplored component of human genomes. We show that HSATII RNA accumulates in...
large nuclear foci in over half of all human cancers examined. This occurs both in vivo and in vitro and is satellite family and locus specific. We also show an unexpected relationship between

Our study extends far beyond the initial observation of HSATII RNA expression to illuminate the fundamental biology in several major ways, giving rise to a concept for the potential significance
of HSATII changes via impact on the epigenome. We examined three satellite families (α-SAT, HSATII, and HSATIII) in an array of tumor types and found preferential expression of HSATII. Zhu et al. (2011) reported α-SAT expression in BRCA1-dependent breast cancers and concluded this was due to BRCA1 loss. However, that study did not examine BRCA1+ tumors for comparison. Our work clarifies that satellite misregulation is widespread in many tumors, irrespective of whether tumors maintain BRCA1, and HSATII RNA was more frequently expressed.

This study provides a conceptual framework to investigate HSATII biology but also demonstrates a powerful approach to study satellites, which can reveal biological relationships that would escape other approaches. Study of such highly repetitive sequences is notoriously difficult by standard genomic approaches. The abundance of these simple repeats makes contamination a nearly intractable problem during extraction (reviewed in Vourc'h and Biamonti, 2011), hence HSATII DNA and RNA hybrids may form in vitro rather than being generated in vivo by a postulated reverse transcriptase (Bersani et al., 2015). Earlier evidence suggested that duplex molecules of highly repetitive sequences were impossible to avoid during RNA extractions (Fedoroff et al., 1977). Similarly, the abundance of HSATII RNA specifically in cancer may also complicate quantification of extracted satellite DNA. Using in situ hybridization, we were able to study the 1q12 satellite (not in the genome build) and found no substantial HSATII amplification within existing loci.

**Biology and Significance of Human Satellite II**

The study of pericentric HSATII has lagged far behind the rest of the genome, but our results provide evidence for its biological significance, based on its capacity to amass and sequester cytological scale (i.e., huge) accumulations of epigenetic factors. Theoretically, if each 26-bp HSATII repeat in the two largest (~6 Mb) 1q12 loci could bind PRC1, just this locus alone could (disproportionately to size) is likely because it contains sequence subtypes not significantly detected on other chromosomes. Similarly, abundant HSATII repeat RNAs aggregated into large nuclear foci can also act as sponges for MeCP2 in cancer nuclei. The effects of HSATII on PRC1 and MeCP2 may be the “tip of the iceberg,” as these and other repeat sequences (half the genome) could potentially impact distribution of many chromatin regulatory factors.

This connection between HSATII DNA methylation and PcG distribution has similarities to findings in normal mouse embryonic stem cells (ESCs) (Cooper et al., 2014), which showed that induced demethylation of mouse major-satellite DNA triggers recruitment of PRC1 and PRC2 to all mouse pericentromeres. PcG factors were also drawn away from canonical gene targets, bolstering the premise here that sequestration of factors on satellite repeats can impact other regions of the epigenome. However, there are interesting differences between our findings in human and those in mouse. In human, PRC1 preferentially localizes to the singular pericentric HSATII locus on 1q12 rather than all pericentromeres as in mice. Also, we generally did not find PRC2 accumulated with PRC1 on satellites as seen in mouse, although our results suggest PRC2 may be required for CAP body formation. Most importantly, findings here show that CAP body formation at 1q12 occurs naturally in vivo, is not a “normal” nuclear structure, and is a hallmark of cancer seen in many tumors.

Given that this relationship to 1q12 is seen in vivo, albeit in a disease context, we postulate that it may have a normal function during development, such as global epigenomic programming in early development, or gametogenesis. Hypomethylation of satellites has been reported to be a hallmark of gametes, preimplantation embryonic cells, and extraembryonic tissues (e.g., Kaneda et al., 2011; Zagradisnik and Kokalj-Vokac, 2000). Similarly, HSATII RNA foci may have some normal role during
Although it remains to be determined whether some satellite indicate a clonal expansion of an early expression profile of the cancer epigenome in different cells of a tumor but rather HSATII RNA foci and particular loci expressed. Such results show homogeneity of phenotype with respect to the number of HSATII RNA positive tumors, cells throughout a given tumor locus specific for a given tumor.

in the nucleoplasm low in PRC1 and why expression appears genesis (Bhatnagar et al., 2014). Stochastic reactivation and expected to result in immediate reactivation, since multiple chromosome, loss of a single repressive histone mark would not be on a broader scale. Based on knowledge of the inactive X chromatin might favor inappropriate growth and neoplastic progression (e.g., a bigger “sponge” for PcG proteins).

HSATII demethylation and consequent CAP body formation are associated with aberrant distribution of ubH2AK119 more broadly. These observations shed light on the unexplained link between global DNA hypomethylation and derepression of specific polycomb target genes (Jin et al., 2008; Reddington et al., 2013) and may also explain why derepression of PcG-targeted genes in ICF cells (or hypomethylated mouse cells) did not correlate with demethylation of gene promoters (as expected) but with dramatic changes in their histone modifications (Jin et al., 2008).

Hence, the collective evidence supports the possibility these changes in HSATII could further compromise heterochromatin on a broader scale. Based on knowledge of the inactive X chromosome, loss of a single repressive histone mark would not be expected to result in immediate reactivation, since multiple chromatin modifications act synergistically to repress chromatin (Csankovszki et al., 2001). Instead, “leaky” maintenance and sporadic activation of different PcG-regulated genes would occur in individual cells. We suggest this would generate a diversity of expression profiles with more opportunity for those that favor neoplastic progression (Carone and Lawrence, 2013; Timp and Feinberg, 2013), as changes in PcG-regulated genes (Wang et al., 2015) and in UbH2A distribution can impact oncogenesis (Bhatnagar et al., 2014). Stochastic reactivation and clonal selection may also explain why not all HSATII loci express in the nucleoplasm low in PRC1 and why expression appears locus specific for a given tumor.

In this context, it is striking to consider that, in a number of HSATII RNA positive tumors, cells throughout a given tumor show homogeneity of phenotype with respect to the number of HSATII RNA foci and particular loci expressed. Such results argue against this being a simple stochastic mismanagement of the cancer epigenome in different cells of a tumor but rather indicate a clonal expansion of an early expression profile. Although it remains to be determined whether some satellite expression can contribute to a selective advantage, we found the chr7 locus expressed in multiple tumors in a small sample, and several studies have reported amplification of chr7 pericentromere is common and may be a neoplastic “driver” in many types of solid tumors (including 80% of metastatic breast cancers) (e.g., Reis-Filho et al., 2006; Sanson et al., 2011). We would also argue that the prevalence of PRC1 “capping” 1q12 in 50% of tumors is clearly non-random, and thus early loss of DNA methylation on HSATII could promote sporadic reactivation of PcG-regulated genes/loci in individual cells, any one of which might favor inappropriate growth and neoplastic progression.

“Toxic Repeat RNAs” in Cancer Can Impact Epigenetic Regulators

This study provides evidence that the HSATII RNA expressed in cancer is not just an inert biomarker of epigenetic dysregulation but can accumulate MeCP2 (and likely other regulatory factors) into aberrant nuclear bodies, possibly contributing to further epigenetic changes. HSATII RNA foci are highly reminiscent of nuclear “toxic repeat RNAs” in certain triplet repeat expansion disorders (myotonic dystrophy type I), in which sequestration of alternative splicing factors by repeat RNA is key to disease pathology (Sicot and Gomes-Pereira, 2013). Sequestration of MeCP2 is unlikely to be innocuous, as MeCP2 is a broad regulator, including miRNA processing (Woo and Kim, 2014) and LINE-1 repression (Yu et al., 2001), which is inappropriately expressed in many cancers (Carreira et al., 2014; Rudić and Burns, 2013). This would explain why we saw increased L1 RNA in our cancer samples and why Ting et al. (2011) noted L1 was overexpressed in cancers expressing HSATII.

Finally, these findings also have potential implications for cancer diagnosis and treatment. First, cancer-specific bodies discovered here may provide biomarkers of epigenetic instability with potential prognostic utility. These bodies can be seen in a single cell of a pathology specimen, and our survey of tumors provides a preliminary indication that this is more common in higher-grade tumors. It is also notable that the demethylating agents used here are in clinical development for cancer therapy (to reactivate tumor suppressor genes) (Kelly et al., 2010); however, our results caution that demethylation of repeat sequences can impact master epigenetic regulators and amplify effects on the nuclear epigenome.

EXPERIMENTAL PROCEDURES

Cell Lines, Growth Conditions, and Fixation

All cell lines (listed in Supplemental Information) were grown in conditions recommended by suppliers (ATCC, Cambrex, and Coriell). 5-aza-2-deoxycytidine (0.2 μg/ml) and 3-deazaneplanocin A (EZH2 inhibitor) (5 μM) were added fresh daily to asynchronously growing cultures. Our standard fixation protocols have been detailed previously (Byron et al., 2013; Johnson et al., 1991; Tam et al., 2002) (details in Supplemental Information).

FISH and IF

Our standard hybridization conditions for RNA, DNA, simultaneous DNA/RNA, and simultaneous DNA or RNA and IF detection was performed as previously described (e.g., Byron et al., 2013) (details in Supplemental Information). Sat2-24nt LNA oligo was used for most images unless otherwise indicated. List of probes are in the Supplemental Information. Oligos were usually hybridized at 15% formamide, unless higher-stringency hybridizations were needed (40% and 90% formamide).
RNA-IP
RIPs were carried out essentially as described (Jones et al., 1998; Long et al., 2011) (details in Supplemental Information).

ChIP-Seq and ChIP-PCR
ChIP-seq and Illumina paired-end deep sequencing were performed as previously described (Yildirim et al., 2011) with some modification (details in Supplemental Information). Deep sequencing data were mapped to human genome build hg19 using Bowtie. Data normalization and peak calling were performed over a 10-kb sliding window using SeqMonk.

Statistics
For the ChIP qPCR data, 95% confidence intervals were calculated (p < 0.05) when comparing between groups (treated and control). Data including error bars of 95% confidence for β-actin gene are shown. ChIP-seq peaks were computed assuming a Poisson distribution and a cutoff of p < 10^-5. Peak overlap defined to be within 50 bp of each summit. p values for disease ontology enrichment were computed using a binomial distribution, corrected for multiple hypothesis testing by the Benjamini and Hochberg procedure.

Microscopy and Digital Imaging
An Axiovert 200 Zeiss microscope was used equipped with a 100 x PlanApo objective (numerical aperture [NA] 1.4) and Chroma 83000 multi-band-pass dichroic and emission filter sets, set up in a wheel to prevent optical shift. Images were captured with the Zeiss AxioVision software and an Orca-ER camera.

ACCESSION NUMBERS
The accession numbers for the UbH2A ChIP-seq data reported in this paper are GEO: GSE94973, GSM2492610, GSM2492611, GSM2492612, and GSM2492613.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, three figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.02.072.

AUTHOR CONTRIBUTIONS
Conceptualization, J.B.L. and L.L.H.; Investigation, M.B., D.M.C., and L.L.H.; Early Experiments, G.P.P.; RNF2 ChIP-seq analysis, T.W.W.; RNA immuno-precipitation, P.J.; Pathology, L.L.H., J.B.L., and D.M.C.; Methodology, J.B.L. and L.L.H.; Visualization, M.B., L.L.H., and D.M.C.; Funding Acquisition, J.B.L.; Supervision, J.B.L.

ACKNOWLEDGMENTS
This work was supported by NIH grants GM107604 and GM053234 supplemented by a grant from the American Cancer Society RG93-033. We thank Dr. Cherie Taglienti from the UMass Tumor Tissue Bank, Heather Kolpa and Alvin Gomez for early FISH experiments, and Takako Jones for help with HSA-til RIP. T.W.W. acknowledges Kunal Rai for helpful correspondence. In 2011, UMass Medical School and authors (L.L.H., M.B., D.M.C., J.B.L.) filed a patent application related to this work, which was published online in 2014.

Received: December 28, 2015
Revised: January 3, 2017
Accepted: February 23, 2017
Published: March 21, 2017

REFERENCES


