Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer

(Article begins on next page)

02 May 2024

Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer

Francesca Bersani,[1] Eunjung Lee,[4,5] Peter V. Kharchenko,[4,6] Andrew W. Xu,[4] Mingzhu Liu,[1,9] Kristina Xega,[1] Olivia C. MacKenzie,[1] Brian W. Brannigan,[1] Ben S. Wittner,[1] Hyunchul Jung,[7] Sridhar Ramaswamy,[1,2] Peter J. Park,[4,5,8] Shyamala Maheswaran,[1,3] David T. Ting,[1,2,10]* Daniel A. Haber,[1,2,9,10]*

[1] Massachusetts General Hospital Cancer Center and Departments of [2] Medicine and [3] Surgery, Harvard Medical School, Charlestown, MA 02129, USA;

[4] Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA;

[5] Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA;

[6] Hematology/Oncology Program, Children's Hospital, Boston, MA 02115, USA;

[7] Samsung Medical Center, Seoul, 135-710, Korea;

[8] Informatics Program, Children's Hospital, Boston, MA 02115, USA;

[9] Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA;

[10] Co-senior authors;

*Correspondence: dhaber@mgh.harvard.edu (D.A.H), dting1@mgh.harvard.edu (D.T.T.)

AUTHOR CONTRIBUTIONS

F.B., D.T.T., S.M. and D.A.H. conceived and designed the study; F.B., K.X., O.M. and B.B. conducted experiments; E.L., P.V.K, A.W.X., H.J and P.J.P. developed analytical tools and performed computational analysis; B.S.W. and S.R. provided bioinformatic support; M.L. supplied technical expertise; F.B. and D.T.T. analyzed and interpreted data; F.B., D.T.T., S.M. and D.A.H. wrote the manuscript. All authors reviewed and revised the manuscript.

ABSTRACT

Aberrant transcription of the HSATII pericentromeric satellite repeat is present in a wide variety of epithelial cancers. In deriving experimental systems to study its deregulation, we observed that HSATII expression is induced in colon cancer cells cultured as xenografts or under non-adherent conditions *in vitro*, but it is rapidly lost in standard 2D cultures. Unexpectedly, physiological induction of endogenous HSATII RNA, as well as introduction of synthetic HSATII transcripts, generated complementary DNA intermediates in the form of DNA:RNA hybrids. Single molecule sequencing of tumor xenografts showed that HSATII RNA-derived DNA (rdDNA) molecules are stably incorporated within pericentromeric loci. Suppression of reverse transcription activity using small molecule inhibitors reduced HSATII copy gain. Analysis of whole genome sequencing data revealed that HSATII copy number gain is a common feature in primary human colon tumors and is associated with a lower overall survival. Together, our observations suggest that cancer-associated derepression of specific repetitive sequences can promote their RNA-driven genomic expansion, with potential implications on pericentromeric architecture.

SIGNIFICANCE

Unique among the large number of non-coding RNA species, the pericentromeric satellite repeat HSATII is massively expressed in a broad set of epithelial cancers, but is nearly undetectable in normal tissues. Here we show that deregulation of HSATII expression is tightly linked to growth under non-adherent conditions, and we uncover an unexpected mechanism by which HSATII RNA-derived DNA (rdDNA) leads to progressive elongation of pericentromeric regions in tumors. The remarkable specificity of HSATII overexpression in cancers, together with the consequences of targeting its reverse transcription, point to a potential novel vulnerability of cancer cells.

\body

INTRODUCTION

Pericentromeric satellite repeats are essential core centromere-building elements that stabilize interactions with DNA-binding proteins, maintain heterochromatin architecture, sustain kinetochore formation, and drive chromosomal segregation during mitosis, thereby ensuring faithful duplication of the genome (1). Transcription from pericentromeric satellites has been reported in plants, invertebrates and during early stages of vertebrate development, and some types of satellite repeats are induced following environmental stress in cell line models (2). We recently reported the massive overexpression of specific classes of satellite repeats in human epithelial cancers, resulting from aberrant transcription of these pericentromeric domains (3). In almost all cancers analyzed, subsets of pericentromeric satellites are expressed at very high levels (3-5), whereas others show consistently reduced expression compared with normal tissues.

The human-specific HSATII satellite is the most differentially expressed satellite subfamily in epithelial cancers (3). It constitutes the main component of pericentromeric heterochromatin on chromosomes 2, 7, 10, 16 and 22 (UCSC Genome Browser, http://genome.ucsc.edu), and it is also found at chromosome band 1q12, where it is colocated with Satellite III sequences. HSATII is defined by tandemly repeated divergent variants of 23-26 bp consensus sequences, organized in long arrays that may span up to thousands of kilobases (6). While repetitive DNA sequences are frequently hypomethylated in cancer cells (7), the mechanisms underlying their aberrant expression are not well understood. For instance, loss of DNA methylation alone does not result in overexpression of the HSATII satellite (8), suggesting the existence of more complex regulatory networks.

In establishing models to study the molecular basis of HSATII RNA overexpression in cancer, we found that growth of cells under non-adherent conditions is sufficient to trigger induction of this satellite repeat. Unexpectedly under these and other conditions, we uncovered that these repeated transcripts are reverse transcribed into DNA:RNA hybrids. The reintegration of HSATII RNA-derived DNA is correlated with a progressive expansion of host HSATII genomic loci. These results point to an unexpected plasticity of pericentromeric repeat-containing structures during cancer progression.

RESULTS AND DISCUSSION

Detection of HSATII expression in human tumors and 3D cancer cell models

The highly repetitive nature of satellites precludes their precise quantitation and qualitative analysis using PCR-based RNA sequencing approaches. We previously showed that PCR-independent single molecule next-generation sequencing (digital gene expression profiling, DGE) is uniquely sensitive and quantitative (9), although it is not suited to routine analysis and does not provide qualitative length information. To enable experimental models for the study of HSATII deregulation, we first designed a modified Northern blot HSATII assay (SI Appendix, Fig. S1A). HSATII satellite transcripts encompass arrays of variable lengths derived from multiple different genomic locations (6), thus Northern blotting generates a pattern of bands ranging from 30 nt to greater than 800 nt in size (SI Appendix, Fig. S1B), consistent with that reported for other repeats, such as murine minor satellites (10) and human Satellite III (11). Quantitation of DGE profiles and Northern blot signal intensity for matched primary cancer specimens were highly correlated (SI Appendix, Fig. S1B and S1C).

We observed that human colorectal cancer cell lines do not express HSATII under standard *in vitro* adherent (2D) culture conditions, but strongly upregulate its expression when grown as tumor xenografts (Fig. 1A). To define specific experimental conditions that modulate HSATII expression within tumors, we tested multiple stimuli associated with cellular stress and tumorigenesis, including hypoxia, UV irradiation, heat shock, oxidative stress, overconfluence, treatment with demethylating agents, co-culturing with stromal-derived feeder layers, and culture under anchorage-free conditions (SI Appendix, Fig. S2A-S2D). Remarkably, only culture under non-adherent conditions, as 3D tumor spheres in solution or in soft agar, led to robust induction of HSATII in five colorectal cancer cell lines, as detected by Northern blotting (Fig. 1B). The specific

induction of HSATII RNA by anchorage-independent growth was also demonstrated using RNA-in situ hybridization (ISH), an imaging assay that does not involve nucleic acid extraction or denaturation, confirming the RNA specificity of the hybridization signal (Fig. 1C). A sixth colorectal cancer line, COLO205, noteworthy for its semi-adherent growth pattern, was unique in expressing HSATII RNA at baseline under standard culture conditions (Fig. 1B). Notably, the elevated RNA levels detected in tumor spheres and xenografts from two independent cell lines were rapidly lost upon re-establishing the growth of these cells under adherent 2D culture (SI Appendix, Fig. S2E). Consistent with our previous findings in primary tumors (3), HSATII transcripts were present in both sense and antisense orientation (SI Appendix, Fig. S2F), and were primarily localized to the nuclear compartment in a panel of colon cancer cell lines (SI Appendix, Fig. S2G), similar to other repetitive ncRNAs such as TERRA (12).

Previous studies have shown that satellite expression in both human and mouse cells may be due to a combination of environmental stimuli and genetic factors. For instance, human Satellite III is expressed in response to cellular stress, including UV-C, oxidative, heat shock, and hyper-osmotic stress (13), while mouse pericentromeric major satellites can be induced by growth to confluence or treatment with hypomethylating, apoptotic, or differentiation-inducing agents (10). Genetic lesions in the tumor suppressor *BRCA1* lead to Alpha satellite sequence derepression in human breast epithelial cells (5) and deletion of *Trp53* in mouse cells results in murine major satellite expression (14). Remarkably, the HSATII pericentromeric repeat is resistant to general environmental stressors and differences in oncogene or tumor suppressor genotypes under standard adherent culture. Only growth under non-adherent conditions is sufficient to induce robust HSATII expression across different cancer cell lines, an interesting phenomenon that merits further investigation.

Identification of medium/small molecular weight HSATII DNA

Unexpectedly, we observed that a fraction of the xenograft-induced HSATII sequences present within medium/small molecular weight nucleic acids (Trizol extraction) was sensitive to DNase I (Fig. 2A). This may result from genomic DNA (gDNA) contamination, or from the existence of small double stranded DNA (dsDNA), double stranded RNA (dsRNA) partially susceptible to DNase I, and/or DNA:RNA hybrids. To first exclude the possibility of gDNA contamination, we performed multiple controls. First during nucleic acid extraction, we applied a solid barrier to prevent any cross contamination between the aqueous phase and the organic, gDNA-containing phase (Phase Lock Gel (PLG) tubes), (SI Appendix, Fig. S3A). Second, we showed that high molecular weight HSATII gDNA from a colon cancer xenograft is readily distinguished by northern blotting from separately processed small molecular weight RNA/DNA from the same tissue, without evidence of cross contamination (SI Appendix, Fig. S3B). Third using multiple cell lines, we showed that the presence of medium/small molecular weight HSATII sequences is only evident following culture under 3D or xenograft conditions. Identical and simultaneous analysis of these cells cultured under 2D conditions showed no evidence of these small HSATII sequences (Fig. 1A, 1B and SI Appendix, Fig. S2). Finally, under identical processing conditions, the rapid loss of HSATII signal following replating of 3D cultures into 2D further excluded gDNA contamination as a source for the HSATII DNA signal on Northern blot (SI Appendix, Fig. S2E). We therefore concluded that deregulated HSATII satellite transcripts coexist with matched DNA fragments within the medium/small molecular weight nucleic acid fraction of cells that overexpress this satellite repeat. Given the presence of small RNA and DNA species, we hypothesized that HSATII RNA may be reverse transcribed into DNA:RNA hybrids and ultimately dsDNA (SI Appendix, Fig. S3C), a phenomenon known to occur with other repetitive elements such as LINEs, SINEs and LTR retrotransposons (15).

Generation of DNA:RNA hybrids upon ectopic introduction of HSATII RNA

To capture the HSATII RNA-to-DNA conversion, we first developed an assay to directly introduce synthetically produced HSATII RNA generated by *in vitro* transcription (IVT) into 2D-cultured 293T cells that lack endogenous HSATII expression (SI Appendix, Fig. S3D). To assess the formation of DNA:RNA hybrids in cells transfected with single stranded IVT HSATII RNA (HSATII-chr10), we subjected nucleic acid extracts to treatment with RNase H, which specifically digests the RNA moiety of DNA:RNA hybrids, but does not affect either single stranded RNA or the DNA component of DNA:RNA hybrids (Fig. 2B). Indeed, RNase H treatment caused a strong reduction in the Northern blot signal identified for the HSATII sense (S; but not anti-sense, AS) sequence (Fig. 2C and S3E). Thus, consistent with the generation of a RNA-derived-DNA (rdDNA), a fraction of the IVT-produced RNA is within a complex with a complementary DNA strand. Transfection of comparable amounts of IVT Green Fluorescent Protein (GFP) RNA produced the expected RNA signal but showed no significant sensitivity to RNase H (Fig. 2D), nor did introduction of GFP RNA lead to the induction of HSATII satellite RNA (Fig. 2C, last two lanes). These findings exclude the occurrence of nonspecific responses to RNA transfection. We conclude that introduction of purified sense HSATII RNA generates a DNA:RNA hybrid in IVT-RNA transfected cells. Since IVT using T7 polymerase relies on a PCR-generated DNA template as starting material, we included multiple controls to ensure the absence of DNA template contamination within the IVT product itself, as well as any genomic HSATII sequences in the cellular extracts (SI Appendix, Fig. S3F-L). These results indicate that ectopically introduced single stranded HSATII RNA is capable of generating complementary DNA within transfected cells.

To further validate these results, we used the S9.6 monoclonal antibody, which is highly specific in its recognition of DNA:RNA hybrids (16-19). We established a

DNA:RNA hybrid immunoprecipitation (DRIP) assay using qPCR of S9.6 immunoprecipitates (HSATII-chr10 qPCR), which was applied to nucleic acids from untransfected or IVT HSATII RNA-transfected cells. Samples were first subjected to complete DNAse I digestion (which removes all dsDNA but does not affect DNA:RNA hybrids), followed by DRIP analysis (Fig. 3A; left). HSATII DNA:RNA duplexes were present only in 293T cells transfected with HSATII RNA and treatment of extracts with RNase H effectively abolished immunoprecipitation of the HSATII DNA:RNA hybrids (Fig. 3B). Taken together, the formation of HSATII RNA:DNA hybrids in a controlled IVT model and the detection of these species by DRIP are suggestive of reverse transcription.

Reverse transcription of endogenous HSATII RNA into DNA:RNA hybrids

To extend these analyses to a more physiological context, we assessed the presence of endogenous HSATII DNA:RNA hybrids by applying the DRIP assay to SW620 colon cancer cells grown as 3D tumor spheres (Fig. 3A; right). RNase H-sensitive DNA:RNA HSATII hybrids were immunoprecipitated using the DRIP assay in SW620 spheroids (Fig. 3C). DNA:RNA hybrids were also identified in COLO205 semi-attached cell cultures (SI Appendix, Fig. S3M), which are characterized by baseline expression of HSATII transcripts (Fig. 1B).

To evaluate the consequences of RT inhibition on the formation of these hybrids, we tested the effect of the nucleoside analog RT inhibitor (NRTI) 2',3'-dideoxycytidine (ddC) in HSATII-expressing cells (Fig. 3A: right). Notably, ddC is very poorly incorporated by replicative polymerases (20, 21), while displaying high specificity for multiple classes of RT, including LINE-1 (22). Treatment of SW620 spheroids and COLO205 cells with ddC significantly reduced the levels of endogenous HSATII DNA:RNA hybrids as measured by the DRIP assay (Fig. 3C and SI Appendix, S3M).

These observations are consistent with RT activity in HSATII-expressing cells, contributing to the generation of DNA:RNA structures derived from satellite transcripts.

RT activity in mammalian cells is derived primarily from retrotransposons, including LINE-1, the HERV family members, and telomerase (hTERT). Definitive identification of the cellular RTs responsible for HSATII reverse transcription is complicated by the diversity of the retrovirally encoded enzyme families and the limited reagents available for their analysis. Attempts to reduce HERV and LINE-1 levels in our cell line models with previously published siRNAs (23) were unsuccessful. Since tools to evaluate hTERT were readily available, we undertook RNA-immunoprecipitation (RIP) analysis of endogenous hTERT from nuclear extracts of SW620 and HCT116 cell lines followed by RT-qPCR for HSATII species. Significant enrichment of HSATII RNA with hTERT immunoprecipitates was evident, relative to control IgG (SI Appendix, Fig. S4A and S4B). Importantly, this enrichment was only observed when the cancer cells were grown as xenografts, but not when they were cultured under standard 2D *in vitro* conditions. As control, the coprecipitation with hTERT of TERC, its primary RNA template for telomere elongation, was unaffected by growth conditions. Negligible coprecipitation with hTERT was observed for an unrelated non-coding RNA target. Furthermore, TERT knockdown using three independent siRNAs demonstrated significant reduction, although not complete depletion, of HSATII rdDNA species (SI Appendix, Fig. S4C and S4D). Thus, hTERT may mediate HSATII reverse transcription, although the contribution of other known cellular RTs cannot be excluded.

Progressive expansion of pericentromeric loci through stable reintegration of HSATII DNA sequences

HSATII-derived rdDNA fragments within the cell nuclear fraction (SI Appendix, Fig. S2G) may either give rise to extrachromosomal elements, or they may be integrated

at chromosomal loci, leading to stable expansion of HSATII genomic sequences. By analogy, RT of LINE-1 transcripts, followed by their retrotransposition at chromosomal loci has been described in epithelial cancers, including colon carcinoma (24). To address this possibility, we first analyzed the dynamics of global HSATII RNA and DNA level changes using single molecule sequencing in SW620 colon cancer cells transitioned from 2D *in vitro* culture conditions to growth as mouse xenografts and vice versa. As expected, the number of HSATII RNA reads was minimal when cells were cultured in 2D conditions, induced 360-fold as the cells gave rise to xenografts in mice, and then promptly downregulated as xenograft-derived tumor cells were returned to *in vitro* 2D cultures (Fig. 4A, blue solid line). Remarkably, total cellular HSATII DNA copy number, which was already abundant at baseline, increased by 25-fold as 2D-cultured cells were transitioned to xenografts (taking into account the multiple length variants of the HSATII tandem repeat unit). The amplified HSATII DNA sequences remained stably expanded despite the downregulation of HSATII RNA transcripts when cells were reestablished under 2D culture conditions *in vitro* (Fig. 4A, red solid line). As a control, we analyzed the GSATII satellite which is structurally similar to HSATII but whose expression is not deregulated in cancer (3). SW620 cells showed negligible GSATII changes, either in RNA or DNA content, as cells transitioned between 2D *in vitro* and xenograft culture conditions (Fig. 4A). Notably, the cancer enriched ALR/Alpha and (CATTC)n satellite RNAs were also induced at high levels in xenografts with conjugate DNA copy number gains, suggesting a common mechanism of RT-mediated genomic expansion of particular repeat classes (SI Appendix, Table S1).

To address the localization of the amplified HSATII DNA sequences, we performed HSATII DNA FISH analysis of 40 xenograft-derived metaphase spreads that did not reveal detectable extrachromosomal elements, nor was any hybridization signal visible outside the five chromosomal loci known to harbor long arrays of pericentromeric

HSATII (Fig. 4B). We could not discern increased fluorescence intensity or size of hybridization signal to demonstrate local expansion due to the limitations of this assay. However, consistent with the FISH data, alignment of HSATII gDNA reads obtained by single molecule sequencing from SW620 xenografts showed that the additional HSATII sequences were distributed among the various endogenous pre-existing HSATII pericentromeric loci (SI Appendix, Fig. S5A and S5B).

To model HSATII DNA copy gain over time and as a function of tumor progression, we serially transplanted SW620 cells as xenografts over multiple generations of mice. Progressive amplification of HSATII gDNA was evident over ten successive rounds of *in vivo* tumor initiation, as measured in cultures derived from xenografted cells. This was assessed using a ~170 bp qPCR-based copy number variation (CNV) assay at the two highest density HSATII pericentromeric regions on chromosome 16q (HSATII-chr16-1 and -2; Fig. 4C-4D and SI Appendix, Fig. S5C-S5D). An adjacent chromosomal region showed no xenograft transplantation-associated copy number changes, ruling out nonspecific gains in the 16q chromosomal arm or in ploidy (Fig. 4E and SI Appendix, Fig. S5E). We did not observe a single chromosome locus duplication event during tumor progression leading to a discrete increase in HSATII genomic DNA. Instead, the pericentromeric genomic loci demonstrated a gradual increase in HSATII gene copy number over time, all within preexisting satellite domains. Such a timeline would be consistent with multiple rdDNA-mediated reintegration events.

Common pericentromeric expansion of HSATII repeats in human colorectal cancers

To determine whether HSATII copy number gains occur in primary human colon cancer, we analyzed CNV in 10 pairs of primary tumors and their matched adjacent normal tissue, focusing again on the chromosome 16q (HSATII-chr16-1 and -2) loci.

After correcting for chr16q arm loss or gain, significantly increased HSATII copy number was evident at either or both of the two independent HSATII loci tested in 5/10 (50%) colon cancers (Fig. 5A-5C and SI Appendix, Fig. S6A). Among other cancers similarly analyzed, HSATII gene copy gain was evident in 5/13 (38%) kidney cancers (SI Appendix, Fig. S6B).

To extend our study of HSATII gene copy changes at specific loci, we performed a genome-wide survey of all such satellite repeats using a novel satellite CNV algorithm to undertake computational analyses of whole genome sequencing (WGS) from The Cancer Genome Atlas Project (TCGA). After correction of these data for large genomic alterations, comparable in size with HSATII stretches, we found that in fully annotated genomic sequences of 51 colorectal cancers (Dataset S1), 23 (45%) had statistically significant genomic gain of HSATII compared with their matched normal germline (Fig. 5D). We extended this analysis to include additional satellite repeats, which revealed higher copy number gains in particular in the satellites whose expression was enriched in cancers (ALR/Alpha, HSATII, (CATTC)n), but not in GSATII and other repeats whose expression is not deregulated in cancer cells (Fig. 5D and SI Appendix, Fig. S6C). HSATII copy gain co-occurred more frequently with ALR/Alpha and (CATTC)n, indicating a common mechanism for repeat expansion in some groups of repeats that is not shared with others (SI Appendix, Fig. S6C). In addition, alignments indicate that repeat expansions occur in the same locations consistent with our xenograft models. Of the 51 TCGA samples, 46 had annotated overall survival data that we analyzed to compare tumors with HSATII gain vs no gain. Notably, Kaplan-Meier analysis demonstrated a significant reduction in overall survival in the HSATII gain vs no gain tumors (Fig. 5E; median OS 1096 vs 1881 days; log-rank p value = 0.034). Taken all together, our data show that gene copy gains at HSATII-encoding pericentromeric repeats and other

cancer enriched satellites occur at pre-existing repeat arrays and are a common and negative prognostic feature of colorectal cancers.

The mechanism underlying HSATII satellite repeat expansion in cancer remains to be defined, but may involve a RT/reintegration phenomenon analogous to that described for other major repetitive elements such as LINE-1 (24) and telomeres (25, 26). While the RT-directed expansion of pericentromeric sequences has not been described in human cells, there is ample precedent for retro-element-mediated integration of such repeats in plants (27, 28). In mammalian cells, integration of non-retroelement sequences within centromeres has not been reported, with the exception of the marsupial tammar wallaby, whose exceptionally short centromeres harbor signatures of retroviral insertions alongside domains of satellite-rich sequences (29). Thus, by analogy with retroviral elements, LINE-1 and telomeres, pericentromeric repeats may expand through the activity of endogenous RT enzymes. However, we cannot unequivocally discriminate between candidate RTs nor can we exclude other mechanisms including replication-induced errors or epigenetic modifier-induced site specific copy gain (30-32).

Suppression of HSATII RT inhibits tumor growth and impairs copy number gains

The critical role of pericentromeric repeats in preserving chromosomal integrity is well established (1). However, the unexpected RT-associated mechanism by which pericentromeres are expanded in cancer cells raises the possibility that its disruption may affect tumor growth. Given the inhibition of HSATII rdDNA formation by treatment of cells with the nucleoside analog ddC, we first tested the effect of this NRTI on cell proliferation. Treatment of SW620 cells with ddC inhibited their proliferation under 3D conditions, but had minimal effect in 2D culture (Fig. 6A and 6B). Mouse tumor xenografts generated from SW620 cells also showed sensitivity to ddC, with a 50% reduction in tumor diameter at 21 days (P=0.03; Fig. 6C). The antiproliferative effect of

ddC was accompanied by a reduction of HSATII copy gain in tumor xenografts (Fig. 6D). In a second colon cancer cell line, HCT116, the growth of tumor xenografts was suppressed using a combination of two RT inhibitors, ddC and d4T, but not using ddC alone, and this effect was again associated with a reduced HSATII copy number gain (SI Appendix, Fig. S7A-S7C). To test an alternative strategy to target HSATII, we synthesized a locked nucleic acid (LNA) oligonucleotide complementary to the HSATII sequence. Treatment of SW620 cells with this HSATII-directed LNA had no effect when the cells were grown in 2D, but it had a strong inhibitory effect on tumor sphere formation (Fig. 6E and F). This effect was associated with accumulation of cells in the G0/G1 phase of the cell cycle (SI Appendix, Fig. S7D). Taken all together, these results raise the intriguing possibility that targeting HSATII transcripts or suppressing cellular reverse transcriptase activity may selectively suppress proliferation of cancer cells under anchorage-independent conditions (SI Appendix, Fig. S7E). Beyond its potential contribution to the viability of proliferating cancer cells proliferating, HSATII transcripts may also play a role in shaping tumor-host interactions as shown in the accompanying manuscript (Tanne et al. PNAS 2015). Thus, the disruption of HSATII RT may have both direct effects on cancer cells as well as modulating the immune response against tumor cells. Given the very high frequency of HSATII deregulation in epithelial cancers (3), such a therapeutic vulnerability might have broad significance.

MATERIALS and METHODS

Human normal and tumor tissues were de-identified discarded excess tissue obtained from the Massachusetts General Hospital (MGH) according to an MGH IRB-approved protocol (2013P001854). The IRB determined consent was not needed for this study. Total RNA from normal human pancreas was purchased from Clontech. The results here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. See SI Appendix for further experimental details.

ACKNOWLEDGMENTS

The authors declare no conflict of interest.

REFERENCES

1. Bierhoff H, Postepska-Igielska A, & Grummt I (2013) Noisy silence: Non-coding RNA and heterochromatin formation at repetitive elements. *Epigenetics : official journal of the DNA Methylation Society* 9(1):53-61.
2. Eymery A, Callanan M, & Vourc'h C (2009) The secret message of heterochromatin: new insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription. *The International journal of developmental biology* 53(2-3):259-268.
3. Ting DT*, et al.* (2011) Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* 331(6017):593-596.
4. Eymery A*, et al.* (2009) A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. *Nucleic acids research* 37(19):6340-6354.
5. Zhu Q*, et al.* (2011) BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* 477(7363):179-184.
6. Jeanpierre M (1994) Human satellites 2 and 3. *Annales de genetique* 37(4):163-171.
7. Ehrlich M (2009) DNA hypomethylation in cancer cells. *Epigenomics* 1(2):239-259.
8. Tilman G*, et al.* (2012) Cancer-linked satellite 2 DNA hypomethylation does not regulate Sat2 non-coding RNA expression and is initiated by heat shock pathway activation. *Epigenetics : official journal of the DNA Methylation Society* 7(8):903-913.
9. Lipson D*, et al.* (2009) Quantification of the yeast transcriptome by single-molecule sequencing. *Nat Biotechnol* 27(7):652-658.
10. Bouzinba-Segard H, Guais A, & Francastel C (2006) Accumulation of small murine minor satellite transcripts leads to impaired centromeric architecture and function. *Proc Natl Acad Sci U S A* 103(23):8709-8714.
11. Rizzi N*, et al.* (2004) Transcriptional activation of a constitutive heterochromatic domain of the human genome in response to heat shock. *Molecular biology of the cell* 15(2):543-551.
12. Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, & Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318(5851):798-801.
13. Valgardsdottir R*, et al.* (2008) Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res* 36(2):423-434.
14. Leonova KI*, et al.* (2013) p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. *Proc Natl Acad Sci U S A* 110(1):E89-98.
15. Levin HL & Moran JV (2011) Dynamic interactions between transposable elements and their hosts. *Nature reviews. Genetics* 12(9):615-627.
16. Boguslawski SJ*, et al.* (1986) Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *Journal of immunological methods* 89(1):123-130.
17. Huertas P & Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell* 12(3):711-721.
18. Hu Z, Zhang A, Storz G, Gottesman S, & Leppla SH (2006) An antibody-based microarray assay for small RNA detection. *Nucleic acids research* 34(7):e52.

19. Rigby RE*, et al.* (2014) RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. *The EMBO journal* 33(6):542-558.
20. Kukhanova M*, et al.* (1995) L- and D-enantiomers of 2',3'-dideoxycytidine 5'-triphosphate analogs as substrates for human DNA polymerases. Implications for the mechanism of toxicity. *The Journal of biological chemistry* 270(39):23055-23059.
21. Louat T*, et al.* (2001) Antitumor activity of 2',3'-dideoxycytidine nucleotide analog against tumors up-regulating DNA polymerase beta. *Molecular pharmacology* 60(3):553-558.
22. Dai L, Huang Q, & Boeke JD (2011) Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC biochemistry* 12:18.
23. Oricchio E*, et al.* (2007) Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene* 26(29):4226-4233.
24. Lee E*, et al.* (2012) Landscape of somatic retrotransposition in human cancers. *Science* 337(6097):967-971.
25. Hastie ND*, et al.* (1990) Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346(6287):866-868.
26. Counter CM*, et al.* (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *The EMBO journal* 11(5):1921-1929.
27. Jiang J, Birchler JA, Parrott WA, & Dawe RK (2003) A molecular view of plant centromeres. *Trends in plant science* 8(12):570-575.
28. Neumann P*, et al.* (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA* 2(1):4.
29. Carone DM*, et al.* (2009) A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. *Chromosoma* 118(1):113-125.
30. Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191(4227):528-535.
31. Cohen Z, Bacharach E, & Lavi S (2006) Mouse major satellite DNA is prone to eccDNA formation via DNA Ligase IV-dependent pathway. *Oncogene* 25(33):4515-4524.
32. Black JC*, et al.* (2013) KDM4A lysine demethylase induces site-specific copy gain and rereplication of regions amplified in tumors. *Cell* 154(3):541-555.

FIGURE LEGENDS

Fig. 1. HSATII is expressed in human tumors and 3D cancer cell models. (*A*) Northern blot analysis of HSATII expression in HCT116 and SW620 cells grown as 2D cultures or xenografts. (*B*) Northern blot analysis of HSATII expression in colon cancer cell lines grown as 2D cultures or tumor spheres (3D). Ethidium bromide (Et Br) stainings of gels are shown for each Northern analysis as loading controls. (*C*) RNA in situ hybridization (with the indicated fluorescent probes) of SW620 cells cultured under 2D conditions or as tumor spheres (3D). HSATII/DAPI colocalization coefficient measured by confocal imaging: *R*=0.6±0.04.

Fig. 2. Ectopic HSATII RNA gives rise to DNA:RNA intermediates. (*A*) Northern blot analysis of HSATII in Untreated (NT) or DNase I-treated extracts obtained from SW620 xenografts. Numbers below indicate relative signal quantitation. (*B*) Trizol extracts obtained from 293T cells 24 h after transfection with HSATII/GFP IVT products were subjected to RNase H treatment followed by Northern blot and hybridization to detect the RNA strand (HSATII S) of the hybrid. (*C*) Northern blot analysis of extracts from 293T cells either untransfected or transfected with IVT HSATII or GFP, subjected to the indicated nuclease treatment and probed for HSATII S. (*D*) Northern blot analysis of extracts from 293T cells after transfection with IVT GFP, treated with RNase H and probed for GFP S. Numbers below indicate relative signal quantitation.

Fig. 3. DNA:RNA hybrid immunoprecipitation (DRIP) reveals the presence of ectopic as well as endogenous HSATII hybrids whose production is affected by reverse transcription inhibition. (*A*) An outline of the experimental layout. Total nucleic acids (TNA) were isolated from IVT HSATII-transfected 293T cells or SW620 tumor spheres

and COLO205 cells cultured in the presence of ddC or DMSO (ddC -). TNA were treated with DNAse I digestion to remove all potential genomic DNA contamination. DNA:RNA hybrids were then purified by immunomagnetic pull down using a hybrid-specific antibody and their relative quantities were measures by HSATII-chr10 qPCR. Pretreatment of TNA samples with RNase H as indicated demonstrates abrogation of DNA:RNA hybrid detection. (*B*) Fold change in the enrichment of DNA:RNA hybrids in HSATII-transfected 293T cells measured by qPCR after DRIP. (*C*) Fold enrichment of endogenous HSATII DNA:RNA hybrids in SW620 tumor spheres analyzed by HSATII-chr10 qPCR after DRIP. Fold changes were calculated based on percent input values and the RNase H-treated samples were set at 1. For all charts, values represent the average of three independent experiments ±SEM. *P*<0.05 (*t* test).

Fig. 4. HSATII rdDNA is re-integrated at the same original loci in the genome, leading to pericentromere elongation in colon cancer xenografts. (*A*) DGE (RNA) and copy number (gDNA) analysis of satellite repeats (HSATII, GSATII) in the indicated samples (SW620) quantitated by single molecule sequencing. (*B*) Representative HSATII DNA FISH (white arrowheads) on metaphase spreads of pre-xenograft 2D cultures and xenografts obtained from SW620 cells. Insets show enlarged HSATII-positive chromosomes. (*C-E*) Copy number variation (CNV) in SW620 cells was assessed by qPCR on (*C*) HSATII-chr16-1 locus, (*D*) HSATII-chr16-2 locus and (*E*) chromosome 16q arm. Ct values for all samples were normalized against β-Actin and DNA CNV is expressed relative to SW620 cells before xenograft implants (pre-xeno), which was set at 1 (T2, T6, T10 = 1 week of culture post-2nd, -6th and -10th serial transplant, respectively). Error bars represent SD (*n*=3).

Fig. 5. Pericentromeric HSATII repeats expand both locally and genome-wide in primary human colon cancer samples. (*A* and *B*) Copy number variation (CNV) analysis of HSATII-chr16-1 (*A*) and HSATII-chr16-2 (*B*) loci on the indicated paired colon specimens (*n*=10). For each sample, values were normalized for β-Actin DNA and corrected for chr16q arm changes. *FC*=mean fold change; *P*=paired *t* test. (*C*) Relative percentage of HSATII copy number changes in colon tumor/normal pairs according to combined HSATII-chr16-1 and -2 CNV analysis, including correction for chr16 arm gains/losses. (*D*) Heat map of whole genome sequencing data on the indicated primary colon cancer specimens based on a log2 ratio cut-off=0.1. (E) Kaplan-Meier curve of overall survival (days) of patients with primary colon cancers with HSATII CNV gain or no gain. *P*=0.034 (log-rank p value).

Fig. 6. RT blockage as well as LNA-mediated inhibition of HSATII transcripts affects tumor sphere growth, impairs tumorigenesis and prevents pericentromeric copy number gains *in vivo.* (*A*) Proliferation assay on DMSO and ddC-treated SW620 cells. Values (%) were normalized against the signal derived from viable cells on the day of seeding, which was set at 100%. (*B*) Tumor sphere forming ability of SW620 cells was tested upon culture in the presence of ddC for 15 days. Values (%) were normalized against the amount of spheres in the DMSO control, which was set at 100%. (*C*) *In vivo* tumor growth of SW620 cell xenografts. Mice were treated daily by IP injection of 25 mg/kg ddC or vehicle alone, starting one week after tumor cell injection. Tumor size at this stage was set at 1 to calculate relative size fold change over time. Error bars represent SEM (*n*=6). (*D*) Copy number variation (CNV) analysis of HSATII by qPCR on chr16-1 locus in tumors recovered from untreated (Vehicle) or ddC-treated mice (*n*=6). Values were normalized for β-Actin and expressed as HSATII/chr16q arm ratios. **P*<0.05 (*t*

test). (*E*) Proliferation assay on SW620 cells upon transfection with a HSATII-specific or Scramble LNA. Values (%) were normalized against the signal derived from viable cells one day after transfection, which was set at 100%. (*F*) Tumor sphere forming ability of SW620 cells was tested upon transfection with a HSATII-specific or Scramble LNA. Values (%) were normalized against the amount of spheres in the control, which was set at 100%. For panels (*A*), (*B*), (*E*) and (*F*) error bars represent SD (*n*=3).