SUPPLEMENTAL INFORMATION

Supplementary Text

Control experiments done for IVT system to rule out DNA template contamination

First, no HSATII band was detected in untransfected cells (Fig. 2C, first lane). Second, no nucleic acid product was visualized by agarose gel electrophoresis unless T7 polymerase was present in the HSATII transcription reaction (IVT; Fig. S3F). Third, only the sense strand (S), corresponding to the transcribed HSATII RNA was evident by Northern blot of the IVT product, and no signal was detectable on the antisense strand (AS; Fig. S3G). Fourth, RNA in-situ hybridization (ISH) analysis of 293T cells (2D culture) transfected with ectopic single stranded IVT sense HSATII RNA showed strong expression of the transcribed sense HSATII sequence with no nonspecific detection of gDNA signal (Fig. S3D). Fifth, presence of the expected 670 nt sense HSATII RNA was readily detected using a single stranded complementary probe in PLG Trizol extracts obtained from transfected cells with no loss of signal following DNase I digestion (Fig. S3H and S3I). Notably, contamination with gDNA sequences would have necessarily resulted in signal reduction upon DNase I treatment. In contrast to the IVT RNA product, when the HSATII-chr10 PCR DNA template itself was directly transfected into 293T cells, signal by Northern blot was not strand-specific and it was completely abrogated by DNase I treatment (Fig. S3K and S3L). Finally, as an additional negative control for specificity of the RNA signal, all control experiments were also performed with IVT GFP (Fig. 2D and S3J). These experiments documented the successful transfection of single stranded sense HSATII RNA into 293T cells, without contaminating template DNA. Together, these results indicate that ectopically introduced single stranded HSATII RNA is unique in generating complementary DNA within transfected cells.

Supplementary Figure Legends

Fig. S1. Optimization of HSATII Northern Blot Assay.

(*A*) An outline of the experimental layout. Derepression of pericentromeric regions in human cancer tissues leads to transcription of HSATII, which is quantified by single molecule Digital Gene Expression (DGE) profiling. In parallel, Northern blotting allows quantitative and qualitative analysis of HSATII levels in tumors and cancer cells subjected to various culture conditions. (*B*) Northern blot analysis of HSATII expression in tissue samples (Normal, normal pancreas; PDAC, pancreatic ductal adenocarcinoma; Duodenal ca., duodenal carcinoma; Well diff NET, well differentiated neuroendocrine tumor). Reads per million (rpm) assessed by DGE profiling are indicated below. (*C*) Linear regression of DGE and Northern blot data (quantified by densitometry) of the same samples.

Fig. S2. HSATII expression in colon cancer cells is not induced by stress but by 3D culture conditions, it is bidirectional and mainly localized in the nucleus.

(*A*) HCT116 cells were grown in low oxygen tension (1%) for the indicated time or UV irradiated for 15 minutes and let recover for 12 hours. RNA was extracted, run on a Northern blot and probed for HSATII. (*B*) Analogous Northern blot on cells non treated (NT) or subjected to either heat shock (1 hour at 42°C followed by increasing recovery periods) or oxidative stress (200 μM H_2O_2 for the indicated times). (*C*) Same Northern blot on either 2D HCT116 cell cultures or the same cells grown to confluence or in medium containing 5 μM 5-azacytidine for 72 h or in soft agar. (*D*) Northern blot of HCT116 RNA extracted from the indicated culture conditions and probed for HSATII. For tumor spheres, co-culture with irradiated MEF feeder layer cells started 5 days after growth as spheres in absence of adhesion. (*E*) Northern blot analysis of HSATII expression in HCT116 and SW620 cells before (2D), during (tum. spheres) and at the

indicated times after (post-TS) culture under non-adherent conditions, or during (xeno) and at the indicated times after (post-xeno) *in vivo* growth. “Post” time points reflect days after replating in 2D conditions. (*F*) HSATII transcript detection by Northern blot in SW620 cells, grown as 2D cultures or xenografts, on both sense (S) and antisense (AS) orientation. (*G*) HSATII Northern blot following differential nuclear/cytoplasmic RNA extraction in 2D cultures, xenografts and tumor spheres (3D) from a panel of colon cancer cell lines. Ethidium bromide (Et Br) stainings are shown for each Northern as loading controls.

Fig. S3. Endogenous and ectopic HSATII signals do not depend on gDNA contamination.

(*A*) Trizol extracts of the same SW620 tumor xenograft, processed in parallel using either regular (REG) or Phase Lock Gel (PLG) tubes, were run on a gel, blotted and probed for HSATII. (*B*) RNA and gDNA extracts from SW620 xenografts were run on a Northern blot. Ethidium bromide (Et Br) staining of gels was used as a loading control. (*C*) Proposed model: HSATII RNA molecules are converted (thick arrow) into dsDNA species through a DNA:RNA hybrid intermediate. Nuclease sensitivity is shown on the left. (*D*) RNA in situ hybridization (with the indicated fluorescent probes) of 293T cells following transfection of IVT HSATII. (*E*) Trizol extracts obtained from 293T cells 24 h after transfection with HSATII IVT product were subjected to RNase H treatment followed by Northern blot and hybridization to detect the complementary DNA strand (HSATII AS) of the hybrid. (*F*) Agarose gel electrophoresis of HSATII-chr10 fragment before (PCR) and after *in vitro* transcription (IVT) with (+T7) or without (-T7) addition of RNA polymerase. (*G*) *In vitro* transcribed HSATII RNA was run on a Northern blot and probed for HSATII on both sense (S) and antisense (AS) strand. (*H*) Extracts from IVT HSATII/GFP-transfected 293T cells were treated with either DNase I or left untreated

(NT) and probed for HSATII S to detect HSATII RNA. (*I*) Northern blot of extracts from 293T cells transfected with IVT HSATII or GFP, subjected to the indicated nuclease treatment and probed for HSATII S. No T7 polymerase (-T7) was added to the control IVT reaction. (*J*) Northern blot of extracts from 293T cells transfected with IVT GFP, treated as indicated and probed with a GFP S-specific oligonucleotide. Ethidium bromide (Et Br) stainings of gels are shown as loading controls. (*K*) HSATII PCR product was transfected in 293T cells, followed by nucleic acid extraction 24 h later. Extracts were either treated with DNase I or left untreated (NT) and probed for HSATII S and AS. (*L*) Northern blot shows no signal on either strand upon DNase I digestion of the extracts obtained as described above. (*M*) Fold enrichment of endogenous HSATII DNA:RNA hybrids in COLO205 cells 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. S4. Pericentromeric HSATII repeats expansion may be associated with a TERT-dependent RT function.

(*A*) RNA-IP (RIP) lysates prepared from SW620 2D cultures or xenografts were immunoprecipitated with hTERT or IgG antibody, reverse transcribed and validated by qPCR for HSATII-chr10, TERC (positive control) and U1 (negative control) RNA enrichment. (*B*) RIP, as described in (*A*) on HCT116 2D cultures and xenografts. (*C*) qPCR analysis of TERT expression 72 h after non-targeting (siNT) or TERT-targeting (siTERT) siRNA transfection in COLO205 cells. (*D*) qPCR analysis on HSATII-chr10 locus following DNAse I (blue chart) or RNase A (red chart) treatment and reverse transcription of extracts obtained from COLO205 harvested 72 h after non-targeting

(siNT) or TERT-targeting (siTERT) siRNA transfection. For all charts, values represent average of three or more independent experiments ±SEM. *P<0.05 (Student's t test).

Fig. S5. HSATII DNA amplification is acquired in cancer cell xenografts.

(*A* and *B*) Schematic representation of the genomic alignments of DNA-sequencing-derived reads, presented as (*A*) fold copy number change and (*B*) absolute reads per million (rpm), in the two indicated experimental conditions (SW620 cells). (*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 as explained in Fig. 4. Statistics were calculated using Pfaffl method. Error bars represent SD (n=3).

Fig. S6. HSATII DNA copy number changes found in primary colon cancer.

(*A*) Average copy number variation (CNV) in colon tumor/normal pairs according to combined HSATII-chr16-1 and -2 qPCR analysis, normalized for β-Actin DNA and corrected for chr16q arm changes. For each pair, HSATII in the normal control was set at 1. Error bars represent SD (n=3). *P<0.05 (t test). (*B*) Pie chart displaying percentages of HSATII copy number changes in kidney tumor/normal pairs according to combined HSATII-chr16-1 and -2 copy number variation (CNV) analysis, including correction for chr16 arm gains/losses. (*C*) Percentage of whole genome amplification in the indicated classes of satellites (first column) across a panel of 51 paired samples (second column), including adjusted P values of copy gains relative to HSATII gain (third column). Co-occurrence of HSATII copy gains and amplification of various satellite repeats is expressed in the last column as relative percentage based on WGS data (out of 23 colon cancer samples showing HSATII expansion compared to their normal counterpart; log2 ratio ≥0.1).

Fig. S7. HSATII DNA amplification can be impaired by reverse transcription inhibition and HSATII inhibition by LNA induces a cell cycle arrest.

(*A*) Proliferation assay on DMSO and NRTI-treated HCT116 cells (ddC, 10 μM; d4T, 10 μM). ddC/d4T combination was chosen based on C-T enrichment in the HSATII sequence. Values (%) were normalized against the signal derived from viable cells on the day of seeding, which was set at 100%. Error bars represent SD (*n*=3). (*B*) *In vivo* tumor growth of HCT116 cell xenografts. Mice were treated with ddC, d4T or combination of the two in the drinking water (approximately 25 mg/kg each), 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*=8). (*C*) Copy number variation (CNV) analysis of HSATII by qPCR on chr16-1 locus in tumors recovered from untreated (Vehicle) or ddC/d4T-treated mice (*n*=3). Values were normalized for β-Actin and expressed as HSATII/chr16q arm ratios. **P*<0.05 (*t* test). (*D*) Cell cycle analysis of SW620 cells grown in tumor sphere medium 48 h after transfection with a HSATII-specific or Scramble LNA. Error bars represent SD (*n*=3). (*E*) Proposed model of HSATII overexpression leading to progressive amplification of pericentromeric loci through a reverse transcription mechanism in 3D cancer cell cultures, xenografts and human tumors. Anchorage-independent culture conditions as well as *in vivo* growth lead to derepression of HSATII pericentromeric repeats in human cancer cells. Once expressed, these sequences undergo reverse transcription that mediates the formation of DNA:RNA hybrids, which eventually give rise to RNA-derived DNA (rdDNA) intermediates. In turn, these molecules display the potential to be reintegrated into the genome within the same pericentromeric regions from which they originally derive. This process gradually induces an expansion of pericentromeric loci in cancer, with implications for the maintenance of genomic integrity and chromosomal stability in tumor cells and important consequences in terms of tumor growth advantage.

Table S1: Comprehensive summary of all satellite repeats identified by DNA- and RNA-seq in the indicated samples derived from SW620 cells

Values are expressed as reads per million (rpm).

Dataset S1: Clinical characteristics and genomic features of TCGA-sequenced colon cancer specimens.

Figure S1

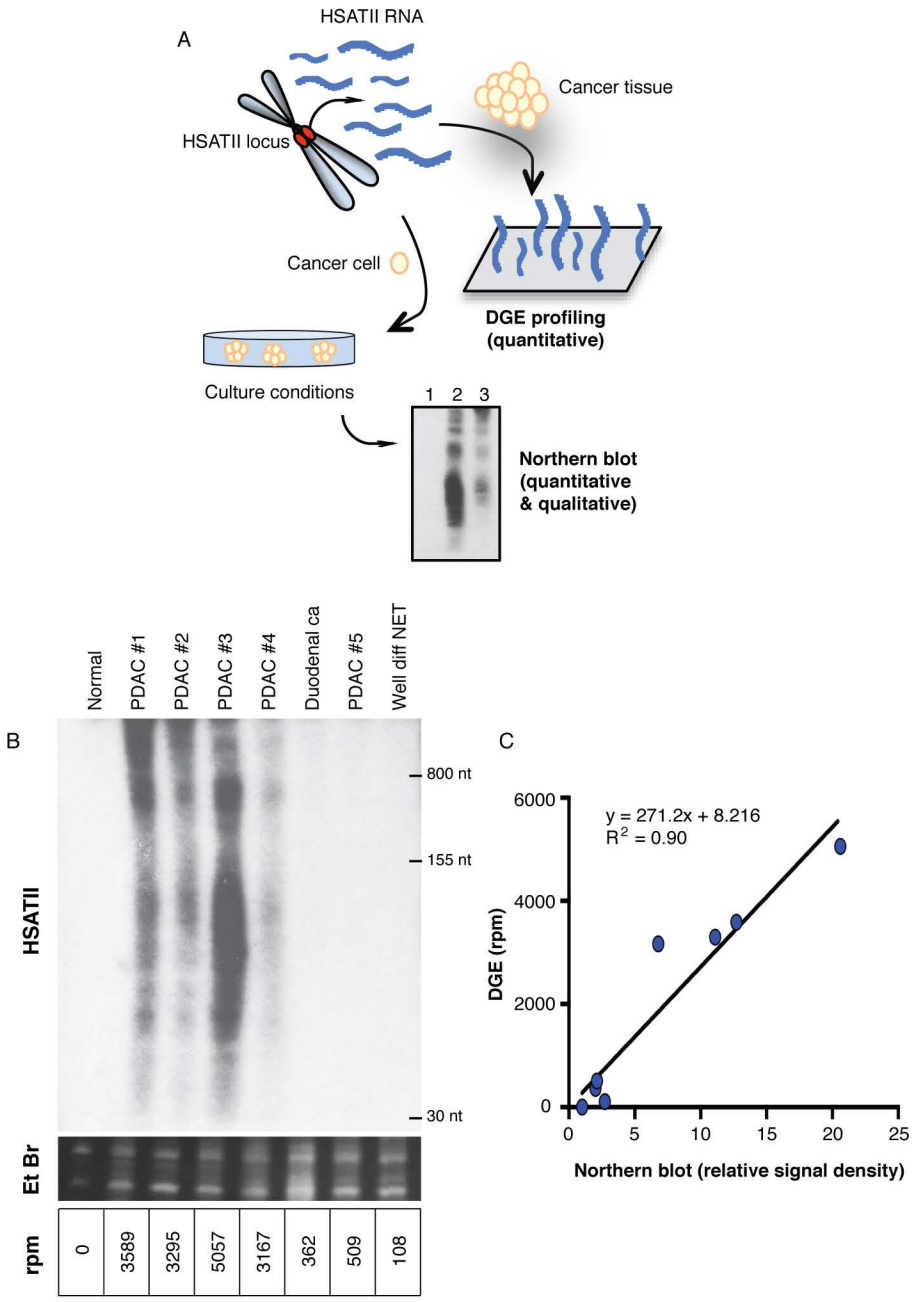

A
HSATII RNA
HSATII locus
Cancer tissue
Cancer cell
DGE profiling
(quantitative)
Culture conditions
1 2 3
Northern blot
(quantitative
& qualitative)
B
Normal
PDAC #1
PDAC #2
PDAC #3
PDAC #4
Duodenal ca
PDAC #5
Well diff NET
HSATII
800 nt
155 nt
30 nt
Et Br
rpm
0
3589
3295
5057
3167
362
509
108
C
y = 271.2x + 8.216
R2 = 0.90
DGE (rpm)
6000
4000
2000
0
0
5
10
15
20
25
Northern blot (relative signal density)

Figure S2

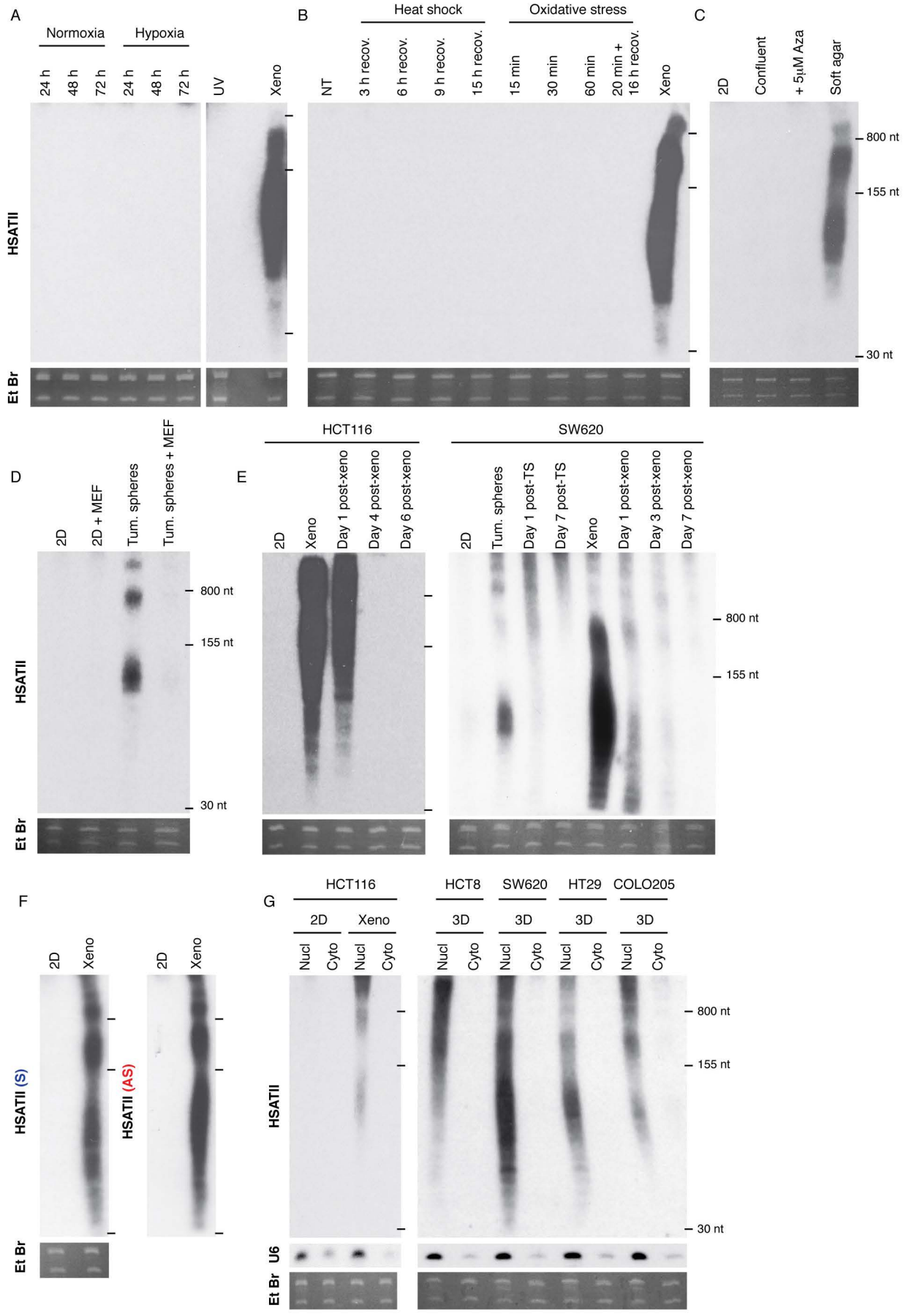

A
Normoxia
Hypoxia
24 h
48 h
72 h
24 h
48 h
72 h
UV
Xeno
HSATII
Et Br
B
Heat shock
Oxidative stress
NT
3 h recov.
6 h recov.
9 h recov.
15 h recov.
15 min
30 min
60 min
20 min + 16 h recov.
Xeno
C
2D
Confluent
+ 5μM Aza
Soft agar
800 nt
155 nt
30 nt
D
2D
2D + MEF
Tum. spheres
Tum. spheres + MEF
800 nt
155 nt
30 nt
E
HCT116
SW620
2D
Xeno
Day 1 post-xeno
Day 4 post-xeno
Day 6 post-xeno
2D
Tum. spheres
Day 1 post-TS
Day 7 post-TS
Xeno
Day 1 post-xeno
Day 3 post-xeno
Day 7 post-xeno
800 nt
155 nt
F
2D
Xeno
2D
Xeno
HSATII (S)
HSATII (AS)
Et Br
G
HCT116
HCT8
SW620
HT29
COLO205
2D
Xeno
3D
3D
3D
3D
Nucl
Cyto
HSATII
U6
Et Br
800 nt
155 nt
30 nt

Figure S3

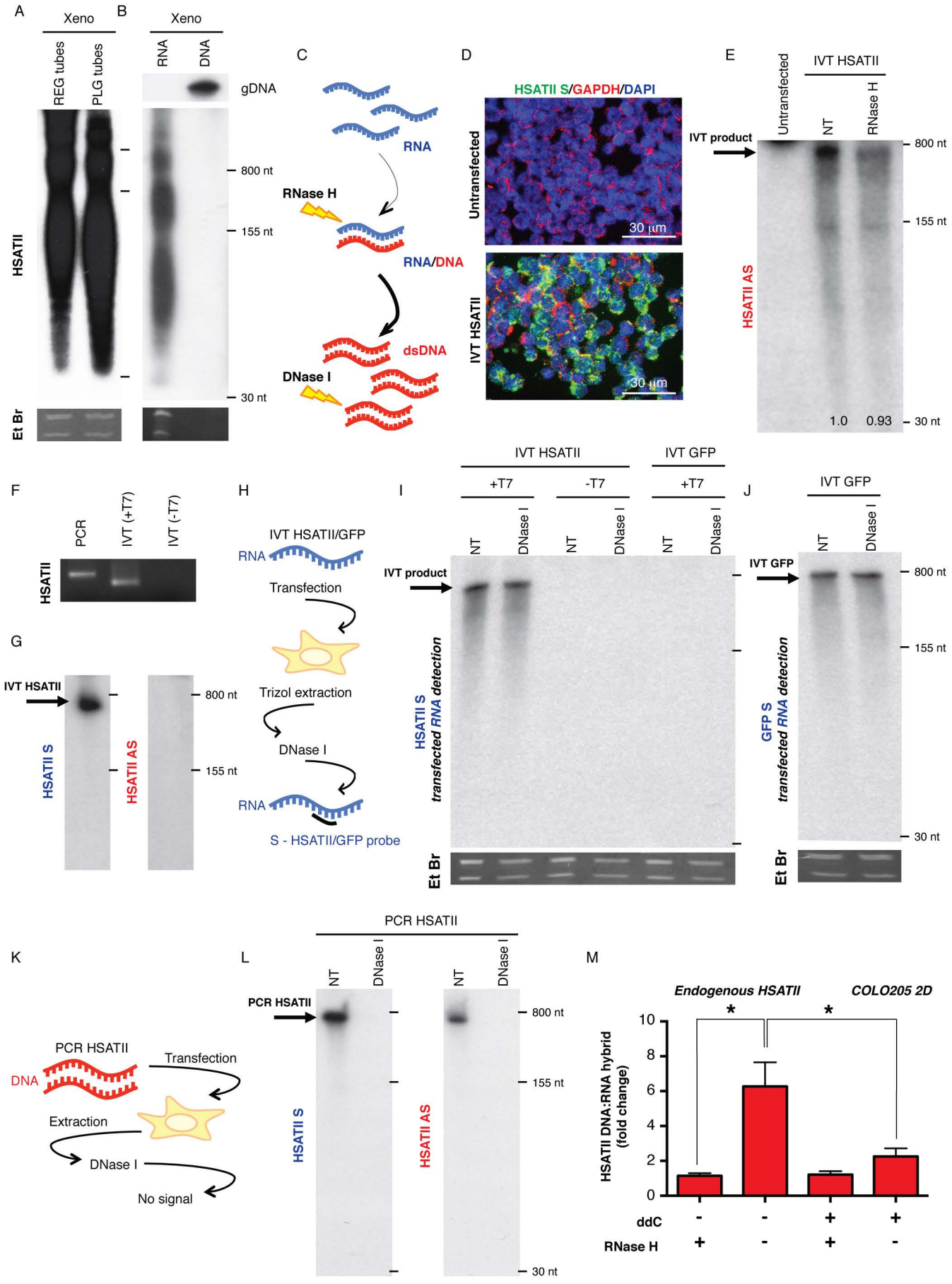

A
Xeno
REG tubes
PLG tubes
HSATII
Et Br
B
Xeno
RNA
DNA
gDNA
800 nt
155 nt
30 nt
C
RNA
RNase H
RNA/DNA
dsDNA
DNase I
D
HSATII S/GAPDH/DAPI
Untransfected
IVT HSATII
30 μm
E
IVT HSATII
Untransfected
NT
RNase H
IVT product
HSATII AS
800 nt
155 nt
1.0
0.93
30 nt
F
PCR
IVT (+T7)
IVT (-T7)
HSATII
G
IVT HSATII
HSATII S
HSATII AS
800 nt
155 nt
H
IVT HSATII/GFP
RNA
Transfection
Trizol extraction
DNase I
RNA
S - HSATII/GFP probe
I
IVT HSATII
IVT GFP
+T7
-T7
+T7
NT
DNase I
IVT product
HSATII S
transfected RNA detection
Et Br
J
IVT GFP
NT
DNase I
IVT GFP
GFP S
transfected RNA detection
800 nt
155 nt
30 nt
Et Br
K
PCR HSATII
DNA
Transfection
Extraction
DNase I
No signal
L
PCR HSATII
NT
DNase I
PCR HSATII
HSATII S
HSATII AS
800 nt
155 nt
30 nt
M
Endogenous HSATII
COLO205 2D
HSATII DNA:RNA hybrid (fold change)
10
8
6
4
2
0
*
*
ddC
RNase H

Figure S4

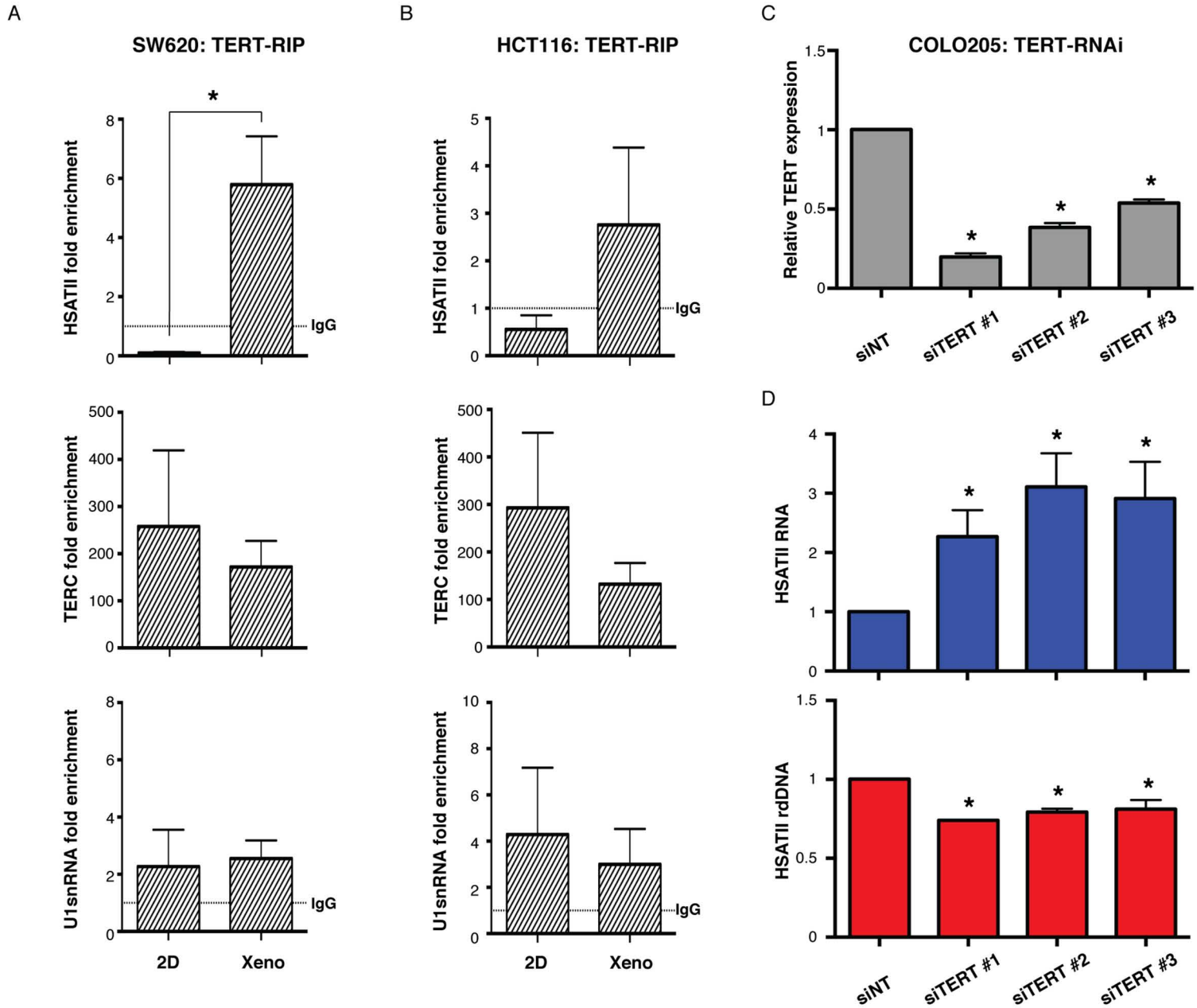
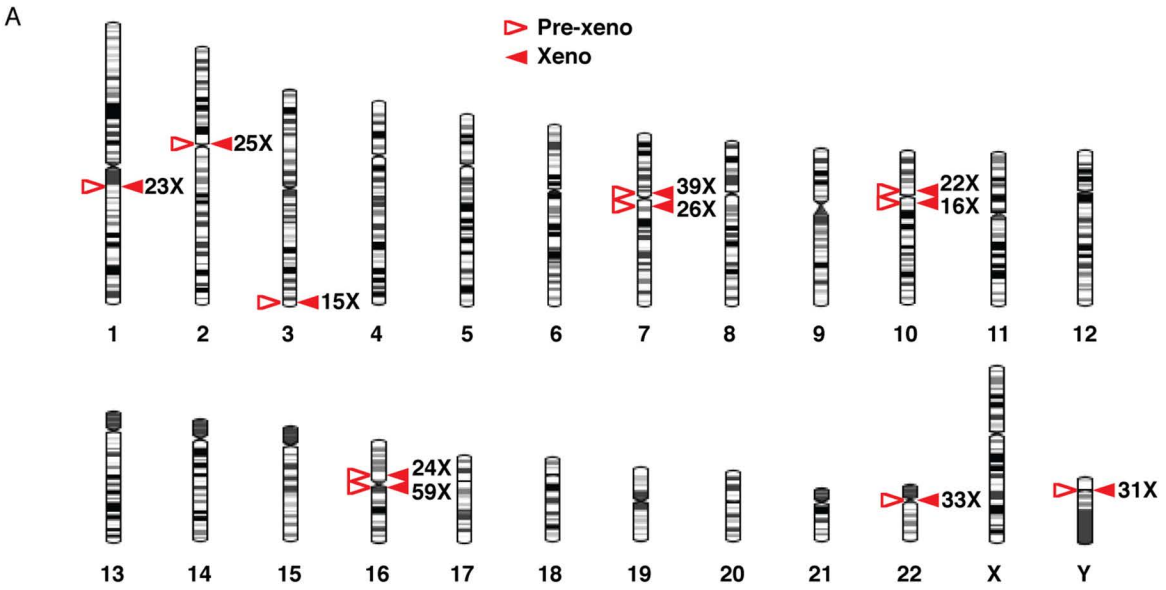
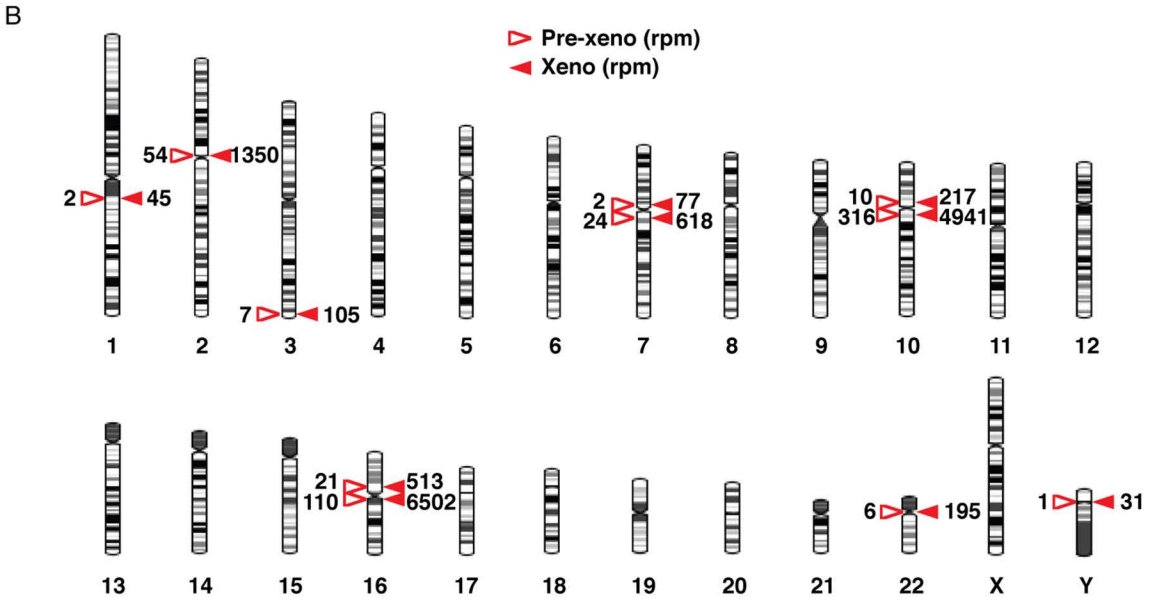
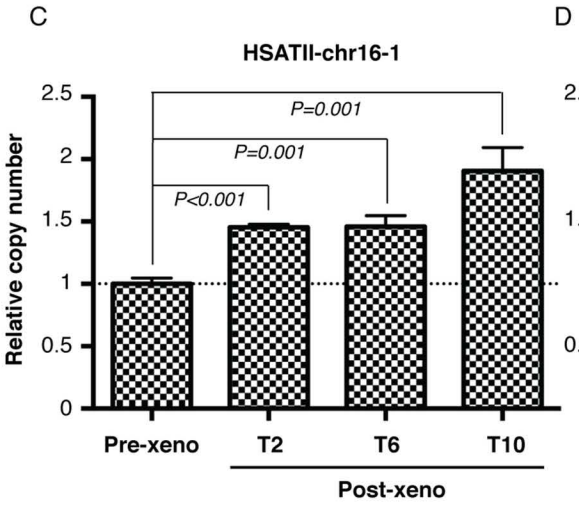
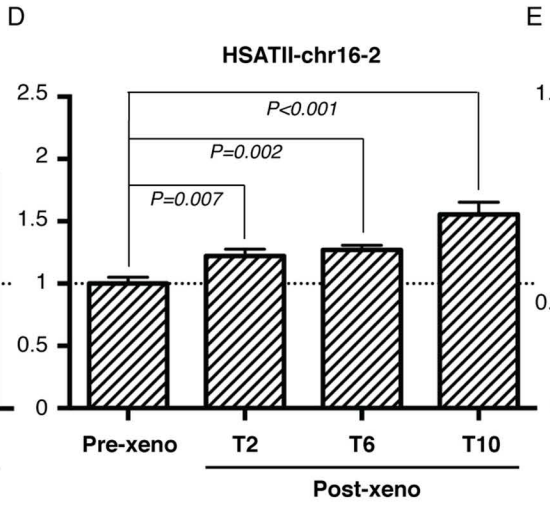

A
SW620: TERT-RIP
HSATII fold enrichment
IgG
TERC fold enrichment
U1snRNA fold enrichment
IgG
2D
Xeno
B
HCT116: TERT-RIP
HSATII fold enrichment
IgG
TERC fold enrichment
U1snRNA fold enrichment
IgG
2D
Xeno
C
COLO205: TERT-RNAi
Relative TERT expression
siNT
siTERT #1
siTERT #2
siTERT #3
D
HSATII RNA
HSATII rdDNA
siNT
siTERT #1
siTERT #2
siTERT #3

Figure S5


A
Pre-xeno
Xeno
23X
25X
15X
39X
26X
22X
16X
24X
59X
33X
31X
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y


B
Pre-xeno (rpm)
Xeno (rpm)
2
45
54
1350
7
105
2
24
77
618
10
316
217
4941
21
110
513
6502
6
195
1
31
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y


C
HSATII-chr16-1
Relative copy number
2.5
2
1.5
1
0.5
0
P=0.001
P=0.001
P<0.001
Pre-xeno
T2
T6
T10
Post-xeno


D
HSATII-chr16-2
2.5
2
1.5
1
0.5
0
P<0.001
P=0.002
P=0.007
Pre-xeno
T2
T6
T10
Post-xeno

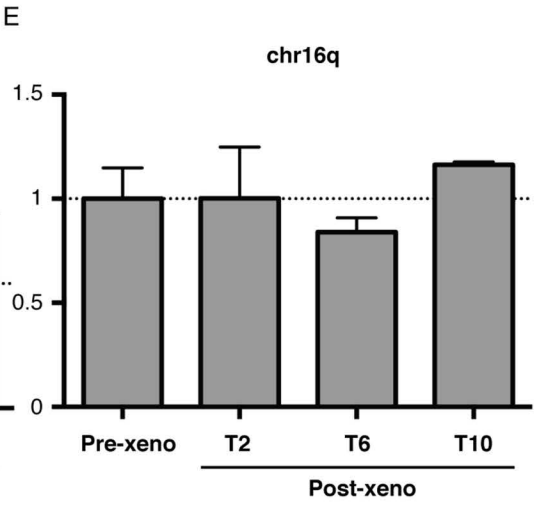
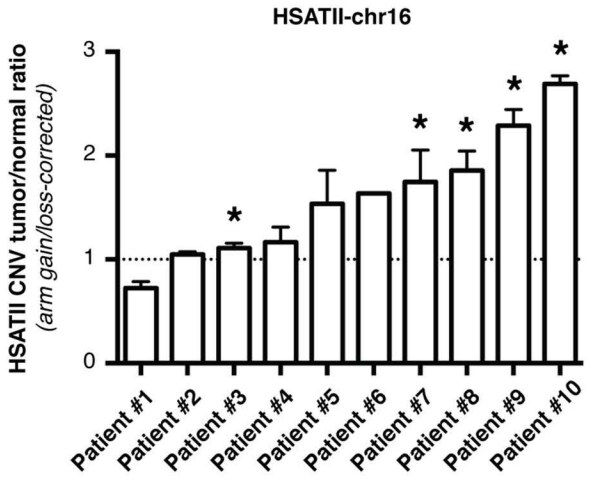
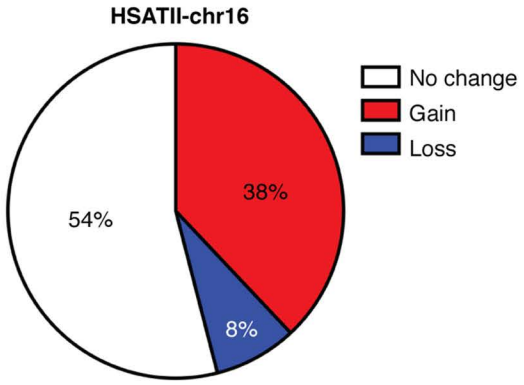

E
chr16q
1.5
1
0.5
0
Pre-xeno
T2
T6
T10
Post-xeno

Figure S6

A



B



C

Satellite repeat	*Copy gain (*)*	*P-value of copy gain difference vs HSATII gain*	*Co-occurence w/ HSATII gain*
HSATII	45%		
ALR/Alpha	51%	6.92E-01	70%
(CATTC)n	29%	1.51E-01	61%
GSATII	16%	2.28E-03	26%
TAR1	14%	9.31E-04	17%
BSR/Beta	4%	6.16E-06	9%

(*) % of samples (51 total) with log2 tumor/normal ratio ≥0.1

Figure S7

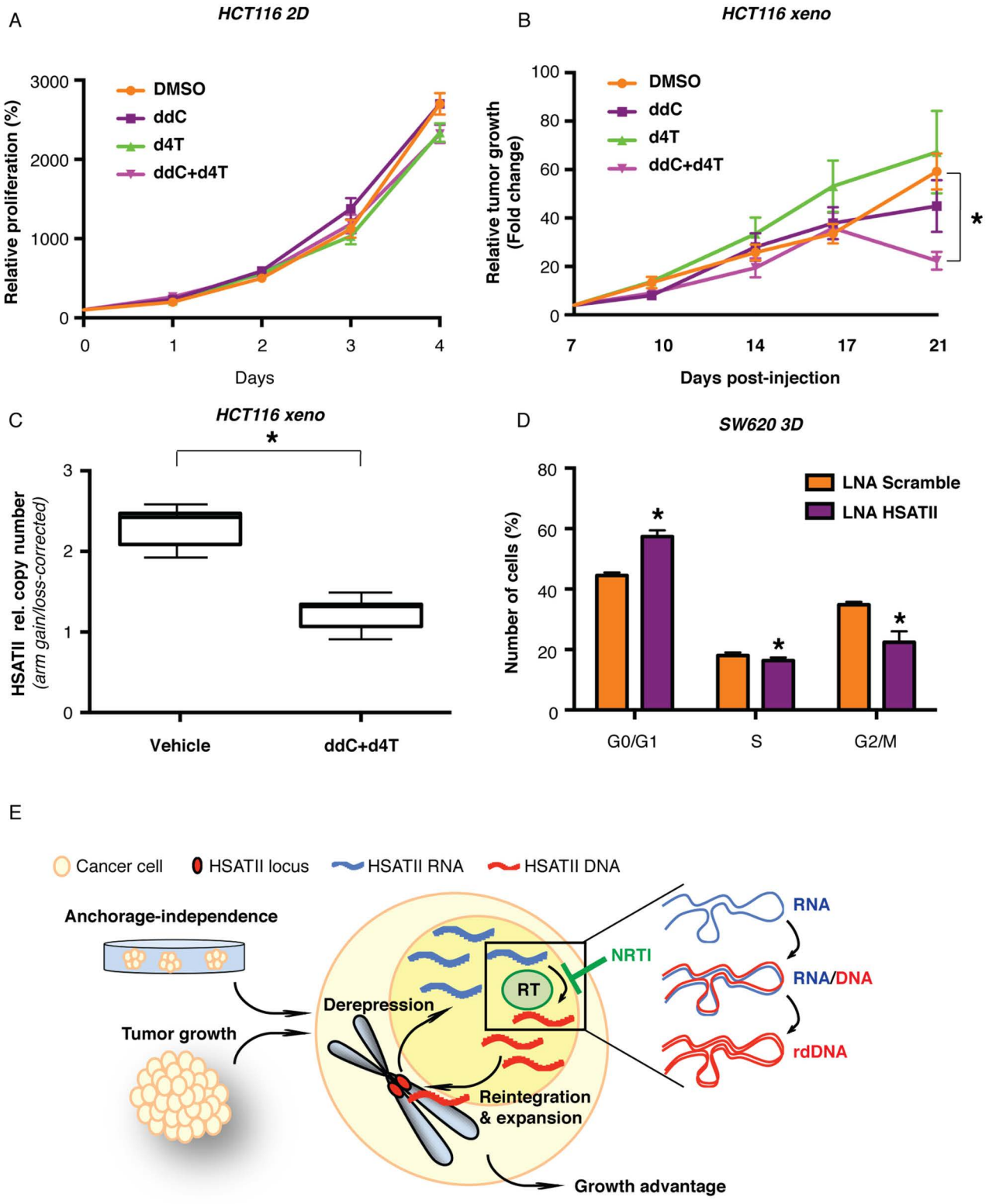

A
HCT116 2D
Relative proliferation (%)
3000
2000
1000
0
DMSO
ddC
d4T
ddC+d4T
0
1
2
3
4
Days
B
HCT116 xeno
Relative tumor growth (Fold change)
100
80
60
40
20
0
DMSO
ddC
d4T
ddC+d4T
*
7
10
14
17
21
Days post-injection
C
HCT116 xeno
*
HSATII rel. copy number (arm gain/loss-corrected)
3
2
1
0
Vehicle
ddC+d4T
D
SW620 3D
Number of cells (%)
80
60
40
20
0
LNA Scramble
LNA HSATII
*
*
*
G0/G1
S
G2/M
E
Cancer cell
HSATII locus
HSATII RNA
HSATII DNA
Anchorage-independence
Tumor growth
Derepression
RT
NRTI
RNA
RNA/DNA
rdDNA
Reintegration & expansion
Growth advantage

Table S1

	DNA CNV (rpm)			*RNA DGE (rpm)*		
Sat. type	Pre-xeno	Xeno	Post-xeno	Pre-xeno	Xeno	Post-xeno
ALR/Alpha	4922	66549	72474	2371	52602	5932
HSATII	518	13066	13143	19	6908	4
BSR/Beta	855	4962	5077	712	3501	1883
(GAATG)n	113	3949	3379	25	2395	37
(CATTC)n	135	3604	3069	19	2044	54
SAR	226	1786	2211	0	782	4
GSATII	545	565	730	305	904	2662
ACRO1	15	1183	707	203	849	158
SST1	168	659	651	254	506	412
SATR1	84	525	595	127	428	62
CER	66	632	588	64	412	37
SATR2	43	361	374	133	267	58
HSATI	3	230	216	0	192	4
REP522	65	194	152	210	294	483
GSAT	44	112	112	127	140	233
TAR1	643	113	104	235	209	187
HSAT4	20	94	83	83	82	150
D20S16	5	66	57	44	46	8
LSAU	16	59	53	19	46	37
GSATX	37	34	38	57	228	604
MSR1	8	45	26	0	28	8
HSAT5	2	37	25	13	32	4
HSAT6	0	9	6	0	5	0
SUBTEL_sa	4	1	2	19	1	4

Experimental Methods

Cell lines and cellular assays

All colon cancer cell lines (ATCC) were cultured in RPMI-1640 + 10% FBS + 1% Pen/Strep (Gibco/Life Technologies). DMEM (Gibco/Life Technologies) with the same supplements was used for 293T cells (ATCC). All cells were tested for mycoplasma contamination giving negative results. For drug treatment and unless otherwise stated, cells were cultured in the presence of either DMSO or 100 μM 2',3'-dideoxycytidine (ddC, Sigma) for two weeks in both 2D and 3D conditions.

For tumor sphere assays, 1 x 10^3 cells were plated as single cell suspension in ultralow attachment 6-well plates (Corning) and grown in serum-free RPMI medium supplemented with 20 μl/ml B27 (Invitrogen/Life Technologies), 20 ng/ml EGF (Invitrogen/Life Technologies) and 20 ng/ml bFGF (Invitrogen/Life Technologies). Fresh medium was added every 3 days. Spheres were collected at day 10 (or at day 15 for ddC experiments). LNA-treated cells were grown as spheres for 5 days, and then quantified by Image J.

For soft agar anchorage-independent growth assays, cells were suspended in 0.4% low-melting agarose (Sigma Aldrich) in 10% RPMI at a density of 2 × 10^4/well, plated on a layer of 0.8% agarose in 10% RPMI in 6-well plates and cultured for 2 weeks.

Xenografts studies and human tissues

Mouse xenograft studies were performed according to an animal protocol approved by the MGH Subcommittee on Research Animal Care. Cells (1 x 10^6) were injected in the flank of six-week old female nude mice (Charles River Laboratories). Tumors were collected 3 weeks later. Cell cultures were reestablished from xenografts after collagenase IV (Sigma) digestion in 1X HBSS (Gibco/Life Technologies) and filtration through a 70 μm cell strainer. For serial transplantation, tumors were minced and a small

piece (~2 mm x 2 mm) was transferred subcutaneously into a recipient mouse. Details about *in vivo* drug treatments are given in the related Fig. legends. 100 mM stock solutions in DMSO were freshly diluted in PBS (Vehicle) before use. Animals did not show any sign of systemic toxicity upon drug administration. Tumors were measured with a caliper every three days and sizes were calculated as the volume of a sphere, then expressed as relative fold changes.

Human normal and tumor tissues were obtained from the Massachusetts General Hospital according to an IRB-approved discarded excess tissue protocol. Total RNA from normal human pancreas was purchased from Clontech.

RNA/DNA extraction and nuclease treatment

Total RNA was extracted with Trizol (Ambion/Life Technologies) according to standard procedures. Where indicated and in all IVT experiments, Phase Lock Gel tubes (5 Prime) were used for Trizol-based extraction. Nuclear/cytoplasmic RNA extracts were obtained using Paris Kit (Ambion/Life Technologies) pursuant to the protocol guidelines. DNA from cells and tissues was extracted with DNeasy Blood & Tissue Kit (Qiagen) according to instructions including the optional RNase A digestion step. Small unintegrated fragments <100 bp were excluded by column purification on a silica-based membrane. Total nucleic acid (TNA) extraction prior to DRIP was done using MasterPure Complete DNA and RNA Purification Kit (Epicentre) according to manufacturer's protocol and followed by complete removal of contaminating DNA by DNase I digestion.

Nuclease treatments (0.5 U/µl DNase I, Roche; 0.5 U/µl RNase H, New England BioLabs; 10 ng/µl RNase A, Ambion/Life Technologies) were carried out following manufacturer's guidelines.

Northern blot

For Northern blot analysis, total RNA (5 µg) before or after nuclease treatment was electrophoresed in a 4% or 8% polyacrylamide-urea gel and transferred by electroblotting onto Hybond-N+ membrane (Amersham/GE Healthcare). Hybridization was performed with the following ^{32}P-labeled DNA oligos: *anti-HSATII S*, 5'-cattcgattccattcgatgat-3'; *anti-HSATII AS* (also indicated in the text and figures as *HSATII*), 5'-atcatcgaatggaatcgaatg-3'; *anti-GFP S*, 5'-cgcgcttctcgttggggtctt-3'; *anti-U6*, 5'-ttgcgtgtcatccttgcgcagg-3'. Relative signal density was quantified with Image J software. HSATII probes were designed on HSATII consensus sequence according to Repbase (http://www.girinst.org) and are specific to the HSATII satellite subfamily (1).

RNA in situ hybridization (ISH)

ISH was performed according to the Affymetrix ViewRNA ISH Cell Assay. Briefly, cells were fixed in 4% formaldehyde for 30 min at room temperature, permeabilized by pretreating in buffer solution for 5 min at room temperature and digested with protease for 10 min. Target probe sets were applied and hybridized to the cells by incubating for 3 h at 40°C. Probes were used at a dilution of 1:50 for Type 1 HSATII and 1:100 for Type 6 GAPDH (Affymetrix). Signal was amplified through the sequential hybridization of PreAmplifier and Amplifier mixes to the target probe set and target RNA molecules were detected by applying Label Probe Mix. Cells were then counterstained with DAPI (5 mg/ml, Life Technologies) and slides were mounted using ProLong Gold Antifade Reagent (Life Technologies). Fluorescence microscopy using a Nikon 90i was used to visualize target mRNAs. Type 1 probes were detected in the Cy3 (550 nm) channel and Type 6 probes in the Cy5 channel (650 nm). Merged images were generated using NIS-Elements software. Colocalization analysis of HSATII and DAPI was performed using

Zen software on 10 single cells (63X magnification) imaged by LSM 710 Confocal Microscope (Zeiss). A Manders overlap coefficient $R \geq 0.6$ indicates signal colocalization.

In vitro transcription and transfection

HSATII was amplified from the genome with the following primers: FOR, 5'-cattccattccattagatgattcca-3'; REV, 5'-tgaatggaataatccttgaacggaa-3' and inserted into pCRII-TOPO vector (Invitrogen/Life Technologies) *via* TOPO-TA Cloning (Invitrogen/Life Technologies). FOR primer is HSATII-specific but maps to multiple genomic locations, whereas REV primer was chosen on a unique sequence to allow amplification of a single HSATII-containing fragment on chromosome 10, which was verified by Sanger sequencing. Notably, the predicted length of the PCR product based on the GRCh37/hg19 genome assembly did not match the actual amplicon size, which reflects the high individual variability, such as for LINE-1 repeats (2), and the low degree of confidence with which pericentromeric regions are currently characterized (1). GFP was subcloned into the BamHI-SalI sites of pSuper vector (Oligoengine). Both fragments were amplified by PCR with M13 for/rev primers and subjected to *in vitro* transcription with T7 RNA polymerase following the T7 transcription kit recommendations (Roche). This included a Turbo DNase digestion step after RNA transcription to prevent DNA carryover. RNA (12.5 µg) was then transfected using Lipofectamine 2000 (Invitrogen/Life Technologies) into a 6 cm tissue culture dish containing sub-confluent 293T cells according to manufacturer's instructions. Cells were collected 24 h post-transfection for nucleic acid extraction.

DNA:RNA hybrid immunoprecipitation (DRIP)

DRIP was performed with minor modifications of a previously published protocol (3). Briefly, total nucleic acid (TNA) extraction followed by complete DNAse I and RNase H

(where indicated) digestion were done as described above. All incubations and wash steps were carried out in 10 mM EDTA/PBS buffer supplemented with RNaseOUT (Invitrogen/Life Technologies). Anti-DNA:RNA Hybrid [S9.6] antibody (ENH001, Kerafast; 12.5 µg/sample) was pre-adsorbed on Protein G Dynabeads (Invitrogen/Life Technology; 30 µl/sample) for 2 h at 4°C. After washing, Protein G-bound antibody was incubated overnight at 4°C with 10 µg of sample, and 1% input was set aside as control. Beads were then washed five times for 10 minutes and immunocomplexes were eluted in 200 µl of 10 mM EDTA/0.6% SDS followed by standard ethanol precipitation. qPCR was performed with the *HSATII-chr10 (IVT)* primer pair for HSATII IVT-transfected 293T cells and the *HSATII-chr10* primer pair for SW620 and COLO205 cells (see primer details below). DRIP fraction Ct values were normalized against input fractions (1%) and fold change was calculated for each sample as the $2^{(-\Delta\Delta Ct)}$ of the normalized "RNase H-" DRIP Ct values over the normalized background ("RNase H+" DRIP fraction), which was set at 1.

qPCR

All qPCR assays were done using Power SYBR-Green PCR Maser Mix on an ABI/PRISM 7500 platform (all reagents were from Applied Biosystems/Life Technologies). The following primers were employed: *HSATII-chr10 (IVT) for*, 5'-caccaaacggaaaaaaacgga-3'; *HSATII-chr10 (IVT) rev*, 5'-gagctcggatccactagtaacg-3'; *HSATII-chr10 for*, 5'-gcattcaattcattagatgacgg-3'; *HSATII-chr10 rev*, 5'-ccttgacccgaatgcaatca-3'; *TERC for*, 5'-tctaaccctaactgagaagggcgt-3'; *TERC rev*, 5'-tgctctagaatgaacggtggaagg-3'; *U1 for*, 5'-caggggagataccatgatcacgaag-3'; *U1 rev*, 5'-ggtcagcacatccggagtgcaatgg-3'. HSATII-chr10 (IVT) primers were designed directly on the validated chr10 cassette contained in the pCRII-TOPO vector used for IVT. HSATII-chr10 primers were designed on a single HSATII locus to generate a unique amplicon

suitable for qPCR purposes. The fragment, verified for HSATII presence by Sanger sequencing, spanned 268 nt rather than 1.6 kb as predicted based on the GRCh37/hg19 genome assembly, reflecting the high individual variability and the low degree of confidence of pericentromeric annotation. TaqMan probes were employed for TERT (Hs00972656_m1) and β-Actin (Hs01060665_g1) analysis on siRNA-transfected COLO205 cells.

For copy number variation assessment on purified DNA, the following primer pairs were used: *HSATII-chr16-1 for*, 5'-ccattcgttaatgcctttcg-3'; *HSATII-chr16-1 rev*, 5'-cacgaatggaatcattgtcg-3'; *HSATII-chr16-2 for*, 5'-tccattcgaggattccactc-3'; *HSATII-chr16-2 rev*, 5'-caaagggaagcaaaggaaatc-3'; *16q for*, 5'-ggggtaaaagggcatgtttt-3'; *16q rev*, 5'-ctgaagaagcccactcaagg-3'; *β-Actin for*, 5'-ctcttccagccttccttcct-3'; *β-Actin rev*, 5'-agcactgtgttggcgtacag-3'. HSATII-chr16 primers were designed based on the GRCh37/hg19 genome alignment and returned specific amplicons of the expected length. For serial transplantation experiments, Ct values for all samples were normalized against β-Actin and DNA CNV was expressed relative to SW620 cells before xenograft implants (pre-xeno), which was set at 1. *E* values for all primer pairs measured by serial dilutions lie close to 2. For this reason, a $2^{(-\Delta\Delta Ct)}$ analysis method was applied. DNA CNV analyzed by Pfaffl method gave consistent results, which are shown for comparison in Fig. S5.

Metaphase spreads and DNA FISH

Cultured cells were trypsinized, resuspended in culture medium and kept in solution in a tube as control for xenograft-derived single cell suspensions. Both preparations were treated with Karyomax Colcemid (Gibco/Life Technologies) at a final concentration of 20 ng/ml. After 4 h of incubation at 37°C, cells were pelleted and resuspended in ice-cold

hypotonic solution (0.56% KCl). After 6 min of incubation at room temperature and a 4 min spin at 1000 rpm, pellets were fixed with methanol:acetic acid (3:1), incubated for 10 min at room temperature, spun again and resuspended in 1 ml fixative. Cells were finally dropped on wet slides from a height of about 45 cm. After "aging" the chromosomes for 72 hours, slides were washed once with 2X SSC for 10 minutes, dehydrated in 70, 90, 100% ethanol for 2 min each and air dried. Probe (Sat.2 Repeat LNA fluorescein probe, Exiqon) was added at a concentration of 1.25 μM in a 50% formamide/2X SSC pH 7.0/10% dextran sulphate solution. DNA was denatured at 75°C for 5 minutes, then hybridized for 4 h at 37°C. After three washes in 0.1X SSC at 50°C for 5 min, two washes in 4X SSC/0.05% Tween at 37°C for 5 min and one wash in PBS at room temperature for 5 min, slides were mounted with Vectashield Mounting Medium with DAPI (Vectorlabs). Pictures were taken at 100X magnification with a Nikon 90i scope with color camera. In total, 40 xenograft-derived metaphase spreads were analyzed.

LNA transfection and cellular assays

500 nM Negative control (Scramble) or HSATII custom designed LNA longRNA GapmeRs (Exiqon) were transfected into SW620 cells with Lipofectamine 2000 (Invitrogen/Life Technologies) following the manufacturer's recommendations. One day post transfection, cells were detached, counted and plated for proliferation or tumor sphere assays. For proliferation assays, cells were seeded at a density of 5 x 10^2 cells/well in a 96-well plate (Corning) and quantified daily using CellTiter-Glo luminescent cell viability assay (Promega) with a SpectraMax M5 microplate reader (Molecular Devices). For tumor sphere assays following LNA transfection, 10^3 cells/well were plated in ultralow attachment 6-well plates (Corning) as described above and collected one or five days later for cell cycle [Propidium Iodide staining (BD Biosciences)

followed by analysis on a BD LSRFortessa flow cytometer (BD Biosciences)] or Image J quantitation, respectively.

RNA-IP (RIP)

At all steps, protease inhibitor cocktail (Roche) and RNaseOUT (Invitrogen/Life Technologies) were added into buffers and solutions. Cells ($1x10^8$) or minced xenografts were crosslinked for 15 minutes with 1% formaldehyde and the reaction was blocked by addition of 0.125 M glycine. After one wash in PBS, pellets were lysed in 6.5 ml of Buffer A (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% NP-40, 0.1 M dithiothreitol) and incubated 30 minutes at 4°C. After centrifuging for 20 minutes at maximum speed at 4°C and a freeze/thaw cycle, we proceeded with protein determination. Equal amounts of lysate (2 mg per antibody) were first pre-cleared with Protein A Dynabeads (Invitrogen/Life Technology) for 2 hours at 4°C and then incubated overnight with 3 µg antibody (Rabbit-anti-TERT, Rockland; Normal rabbit IgG, Cell Signaling) at 4°C. 1% input was set aside before addition of antibodies. Fresh Dynabeads (40 µl) were added to the lysate and incubated 2 hours at 4°C. After binding, beads were washed three times for 30 minutes with Buffer A and immunocomplexes were eluted in 250 µl Elution buffer (100 mM NaHCO3, 1% SDS). Samples were then de-crosslinked by addition of 200 mM NaCl and incubation at 65°C for 2 hours. RNA extraction was performed after addition of 750 ul TRIzol-LS (Ambion/Life Technologies) according to standard procedures and was followed by DNase I digestion and reverse transcription. After qPCR (see details above), RIP fraction Ct values were normalized against input RNA fractions (1%) and fold enrichment was calculated for each sample as the $2^{(-\Delta\Delta Ct)}$ of the normalized TERT-RIP Ct values over the normalized background (IgG-RIP fraction), which was set at 1. Mean values shown in Figure S4A and S4B are from biological replicates of 4 2D and 10 xenograft samples.

siRNA transfection

50nM ON-TARGETplus Non-targeting pool siRNA (Thermo Scientific) and Mission Pre-designed TERT siRNAs (Sigma Aldrich) were electroporated into COLO205 cells according to the Neon Transfection System (Invitrogen/Life Technology) instructions using the following conditions: 1600V; pulse width 10; 3 pulses.

Single molecule sequencing and data analysis

Purified RNA was subjected to digital gene expression (DGE) sample prepping and analysis on the HeliScope Single Molecule Sequencer from formerly Helicos BioSciences and now SeqLL. This method has been previously described (4). Briefly, single stranded cDNA was reverse transcribed from RNA with a dTU25V primer (a modified version of oligo-dT priming) and the Superscript III cDNA synthesis kit (Invitrogen/Life Technologies). Purified single stranded cDNA was denatured and then a poly-A tail was added to the 3' end using terminal transferase (New England Biolabs).

Purified DNA was subjected to DNA sequencing sample prepping protocol from Helicos as previously described (5). Briefly, genomic DNA was sheared with a Covaris S2 acoustic sonicator producing fragments averaging 200 bps and ranging from 100 to 500 bps. Cleaned-up DNA was then denatured and a poly-A tail was added to the 3' end using terminal transferase (New England Biolabs).

Tailed cDNA or DNA were then hybridized to the sequencing flow cell followed by "Fill and Lock" and single molecule sequencing. Sequence reads were subjected to filtering for a minimum read length of 25 and removal of artifact reads followed by alignment of reads to the human genome (UCSC hg19) as well as the satellite reference library from Repeat Masker (6) using the indexDPgenomic aligner (4). RNA and genomic DNA sequence reads aligning to satellites were then normalized to total genomic alignments

to obtain reads per million (rpm). Notably, oligo-dT-based cDNA library construction demonstrated superior activity in HSATII transcript coverage compared to a previously tested random hexamer-based method (~4.5 fold difference), suggesting that HSATII species are characterized by the presence of a polyA tail, a required feature for LINE-1 RT targets. Other classes of repeats, such as GSATII, did not show such variation. All single molecule sequencing data has been uploaded to NCBI GEO ID GSE64471.

Estimating satellite copy-number changes from whole genome sequencing

We downloaded a total of 51 pairs (tumor and matched tissue or blood samples) of colorectal whole genome sequencing (WGS) data from The Cancer Genomics Hub (CGHub; https://cghub.ucsc.edu/). All 51 pairs were sequenced by The Cancer Genome Atlas (TCGA) project. To detect amplification or deletion of satellite DNA in the high-coverage (> 30X) cancer WGS data, we developed a computational method, called *repeat-CNV* (repeat copy number variation; *manuscript in preparation*). Briefly, we built a repeat sequence library that consists of consensus sequences of satellite repeats from Repbase (http://www.girinst.org/repbase/) and satellite repeat sequences annotated in the human genome reference by RepeatMasker (http://www.repeatmasker.org/) (7). All of the sequencing reads were mapped to the custom-built repeat sequence library using a revised *bwa* algorithm to assure that each read was uniquely associated with one satellite repeat (7). The degree of enrichment and depletion of the satellite repeats were estimated based on the ratio of total read counts uniquely mapped to the satellite repeat in a cancer genome versus its matched normal (tissue or blood) genome using the previously described statistics (7). The read counts were adjusted for the GC-bias estimated from reads mapped to genomic copy-neutral regions and for broad copy number change, such as aneuploidy (10 Mb copy number correction was applied). An improved version of BIC-seq (8), called *Integer* (*manuscript in review*) was used to

estimate copy number as well as purity and ploidy of cancer genomes.

Percent copy gain for satellite repeats was based on a threshold of log2 tumor/normal ratio ≥0.1 and statistical significance for differences in copy gains (HSATII versus other repeats) was calculated using the Fisher Exact test with correction for multiple hypothesis with the Holm method.

Kaplan Meier Survival Analysis

Overall survival data was obtained from the UCSC Cancer Browser from TCGA_COADREAD version 2015-02-24 (https://genome-cancer.ucsc.edu). Plots and log rank p value was calculated using GraphPad Prism software. Of the 51 TCGA samples analyzed, 46 had usable annotated overall survival data where 4 of 5 had no applicable data and the other patient appeared to have not survived post-surgery (death at day 0).

Statistics

All *P* values were calculated based on a 2-tailed *t* test with GraphPad Prism 5 software. All sets of data met normal distribution and displayed homogenous variance. Values of $P<0.05$ were considered significant. The sample size of each experiment is specified in Fig. legends. Error bars represent SD or SEM for technical or biological replicates respectively (see Fig. legends for details).

Supplemental References

1. Altemose N, Miga KH, Maggioni M, & Willard HF (2014) Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLoS computational biology* 10(5):e1003628.
2. Ewing AD & Kazazian HH, Jr. (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res* 20(9):1262-1270.

3. Rigby RE*, et al.* (2014) RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. *The EMBO journal* 33(6):542-558.
4. Lipson D*, et al.* (2009) Quantification of the yeast transcriptome by single-molecule sequencing. *Nat Biotechnol* 27(7):652-658.
5. Pushkarev D, Neff NF, & Quake SR (2009) Single-molecule sequencing of an individual human genome. *Nature biotechnology* 27(9):847-850.
6. Ting DT*, et al.* (2011) Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* 331(6017):593-596.
7. Day DS, Luquette LJ, Park PJ, & Kharchenko PV (2010) Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome biology* 11(6):R69.
8. Xi R*, et al.* (2011) Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. *Proceedings of the National Academy of Sciences of the United States of America* 108(46):E1128-1136.