iScience, Volume *23*

Supplemental Information

Polyadenylation of Histone H3.1 mRNA Promotes Cell Transformation by Displacing H3.3 from Gene Regulatory Elements

Danqi Chen, Qiao Yi Chen, Zhenjia Wang, Yusha Zhu, Thomas Kluz, Wuwei Tan, Jinquan Li, Feng Wu, Lei Fang, Xiaoru Zhang, Rongquan He, Steven Shen, Hong Sun, Chongzhi Zang, Chunyuan Jin, and Max Costa

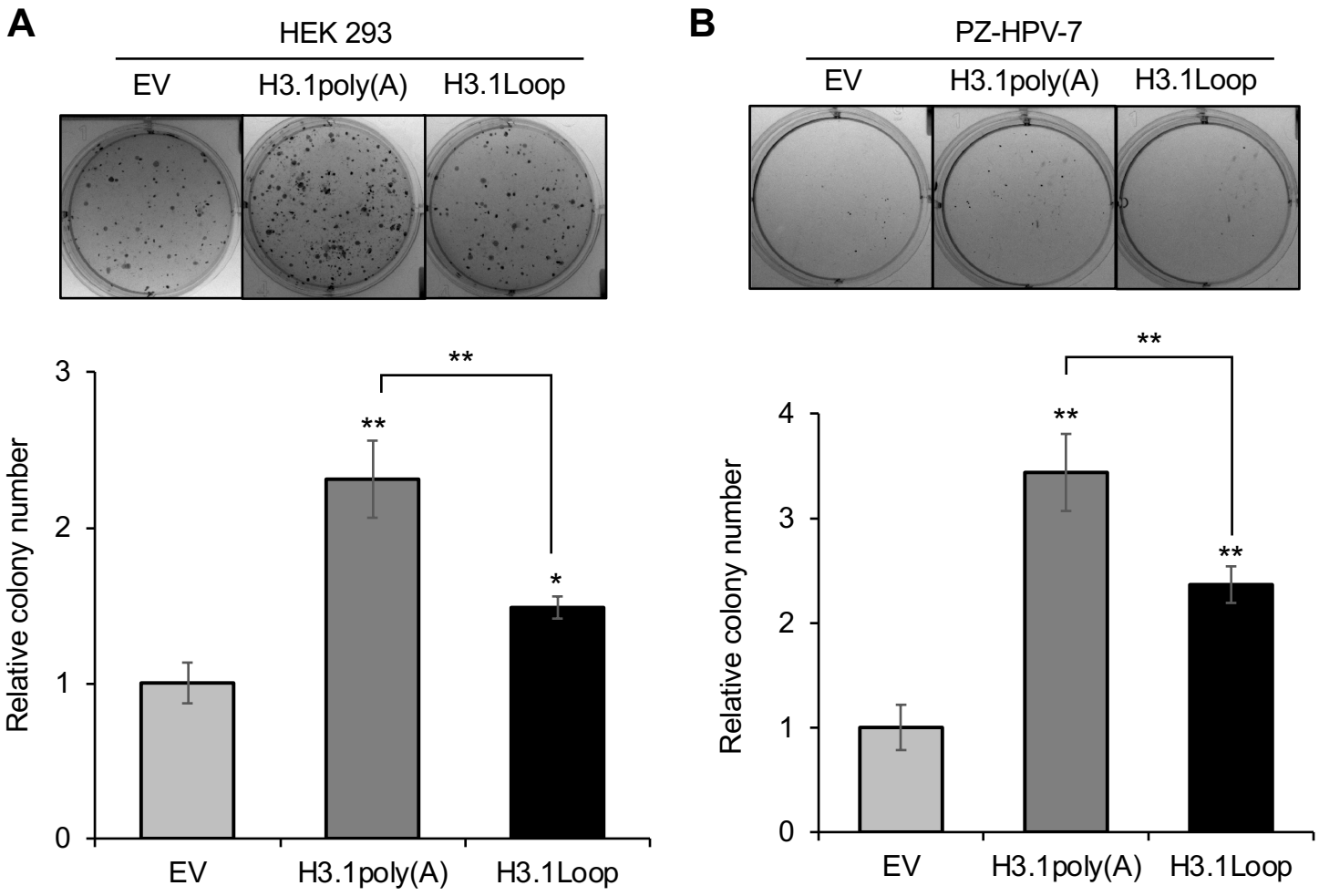

Figure S1, related to Figure 2. Polyadenylation of canonical histone H3.1 mRNA enhances anchorage-independent growth of HEK 293 and PZ-HPV-7 cells

Soft agar assays were performed with HEK 293 cells **(A)** and PZ-HPV-7 cells **(B)** transiently transfected with the empty vector (EV), H3.1poly(A) or H3.1Loop. Approximately 5,000 cells were resuspended and seeded in the media with 0.35% agar over a 0.5% agar base layer in each well of a 6-well plate. After a two- (HEK 293) or four-week (PZ-HPV-7) culture, the colonies were stained by INT/BCIP solution and quantified by the ImageJ software. The number of colonies were adjusted with plating efficiency and adjusted number for the EV control was set to 1. The data shown are the mean ±S.D. (n=3). Student's t test was applied for statistical significance: * $p<0.05$; ** $p<0.01$.

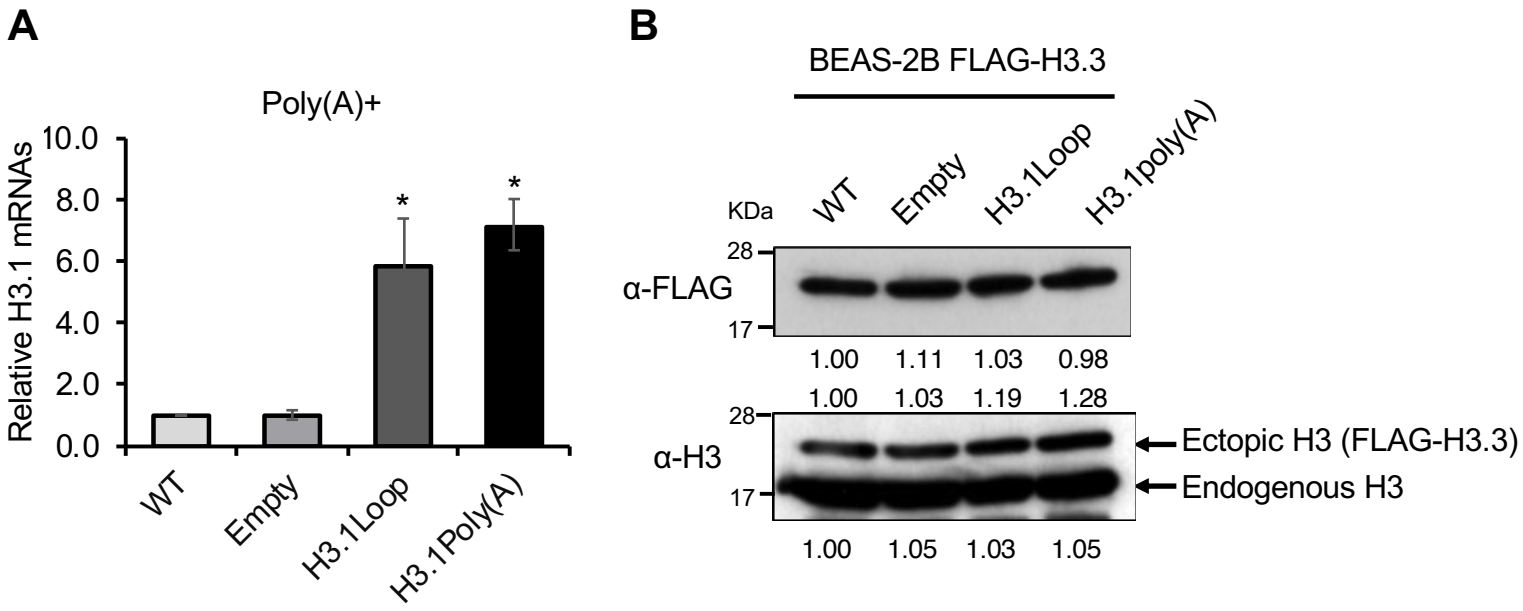

Figure S2, related to Figure 3. The level of FLAG-tagged H3.3 is not changed by ectopic expression of H3.1poly(A)

(A and B) pcDNA-Empty (EV), pcDNA-H3.1Loop (H3.1Loop), or pcDNA-H3.1Poly(A) (H3.1poly(A)) vector was stably transfected into BEAS-2B FLAG-H3.3 cells separately. (A) RT-qPCR results. Total RNA was extracted from each cell lines. mRNA was converted to cDNA using oligo dT primers. Polyadenylated H3.1 mRNA levels were then measured by quantitative PCR. Relative mRNA levels were normalized to *Actin* as an internal control. The data shown are the mean ±S.D. (n=3). Student's t test was applied for statistical significance: * $p<0.05$. (B) Western blot results. Ectopic expression of polyadenylated H3.1 mRNA did not affect ectopic expression of FLAG-H3.3.

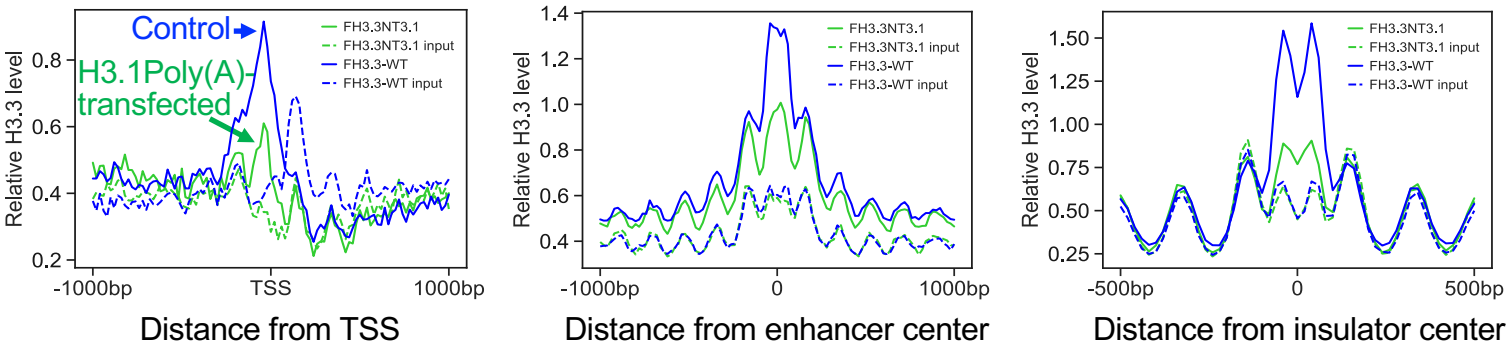

Figure S3, related to Figure 3. Overexpression of polyadenylated H3.1 mRNA displaces histone variant H3.3 from critical gene regulatory elements

Profile of FLAG-H3.3-containg nucleosomes across the transcription start sites (TSSs) for 2,000 highly active genes (activate promoters) (left), DNase I hypersensitive sites (enhancers) (middle), or CTCF-binding sties (insulators) (right) are shown. ChIP signal (solid line) and Input signal (dot line) are presented separately. In the control cells (blue), FLAG-H3.3 was enriched at active TSSs, enhancers, and insulators, respectively. The levels were greatly reduced by the ectopic expression of polyadenylated H3.1 mRNA (green).

A

Diseases and Disorders	p-value range	# Molecules
Cancer	1.41E-05 - 1.62E-15	301
Organismal Injury and Abnormalities	1.41E-05 - 1.62E-15	301
Endocrine System Disorders	1.41E-05 - 6.17E-12	245
Dermatological Diseases and Conditions	1.41E-05 - 1.38E-08	8
Hematological Disease	1.41E-05 - 1.38E-08	14

B Top Upstream Regulators

Name	p-value
MAPT	7.92E-09
RB1	3.22E-08
torin1	2.12E-07
TP53	2.54E-07
ADRB	9.55E-07

C Top Canonical Pathway

Name	p-value	Overlap	
Mitochondrial Dysfunction	1.92E-08	8.8 %	15/171
Sirtuin Signaling Pathway	1.65E-07	6.2 %	18/291
Oxidative Phosphorylation	3.75E-07	10.1 %	11/109
Remodeling of Epithelial Adherens Junctions	4.71E-06	11.8 %	8/68
Phagosome Maturation	5.39E-05	6.6 %	10/151

D Top Networks

ID	Associated Network Functions	Score
1	Carbohydrate Metabolism, Lipid Metabolism, Small Molecule Biochemistry	61
2	RNA Post-Transcriptional Modification, Cardiovascular System Development and Function, Cellular Movement	50
3	Post-Translational Modification, Protein Degradation, Protein Synthesis	45
4	Developmental Disorder, Hereditary Disorder, Metabolic Disease	42
5	Small Molecule Biochemistry, Lipid Metabolism, Nucleic Acid Metabolism	40

Figure S4, related to Figure 4. RNA-Seq analysis of H3.1poly(A) down-regulated genes

(A) Illustrations of top diseases and disorders associated with 328 poly(A)-specific down-regulated genes.
(B) Illustrations of top upstream regulators associated with 328 poly(A)-specific down-regulated genes.
(C) Illustration of top canonical pathway on 328 poly(A)-specific down-regulated genes.
(D) Illustration of top networks associated with 328 poly(A)-specific down-regulated genes.

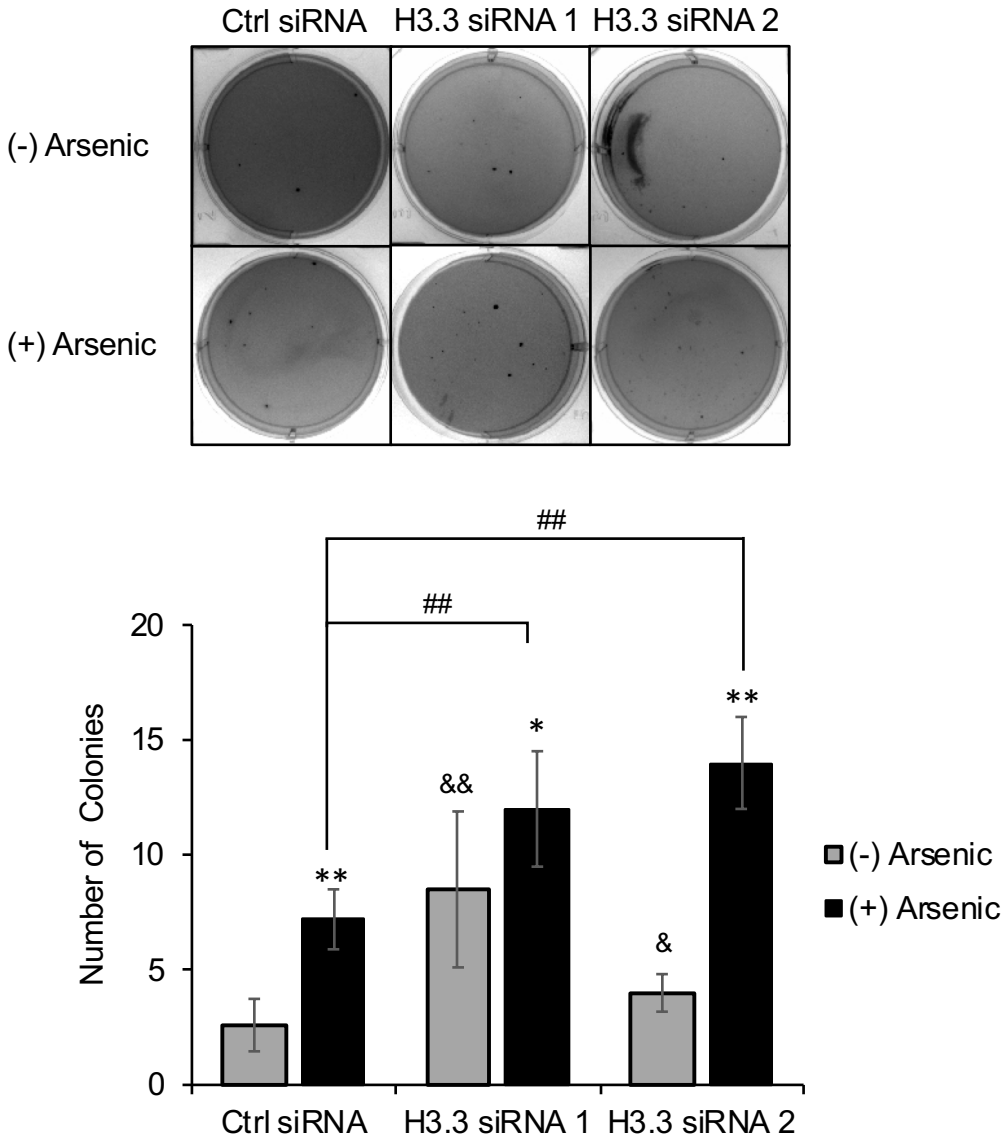

Figure S5, related to Figure 6. Knockdown of H3.3 promotes arsenic-induced anchorage-independent cell growth

BEAS-2B cells were transfected with control (Ctrl) siRNA or with two distinct siRNAs for H3.3 in the presence or absence of 1 μM arsenic. Ninety-six hours after the arsenic treatment, the cells were plated in 0.35% soft agar and cultured for 4 weeks. The data shown are the mean ±S.D. from experiments performed in triplicate. Student's t test was applied for statistical significance: * $p<0.05$ (+) arsenic vs. (-) arsenic, ** $p<0.01$; & $p<0.05$ H3.3 siRNA vs. Ctrl siRNA without arsenic, && $p<0.01$; and ## $p<0.01$ H3.3 siRNA vs. Ctrl siRNA with arsenic.

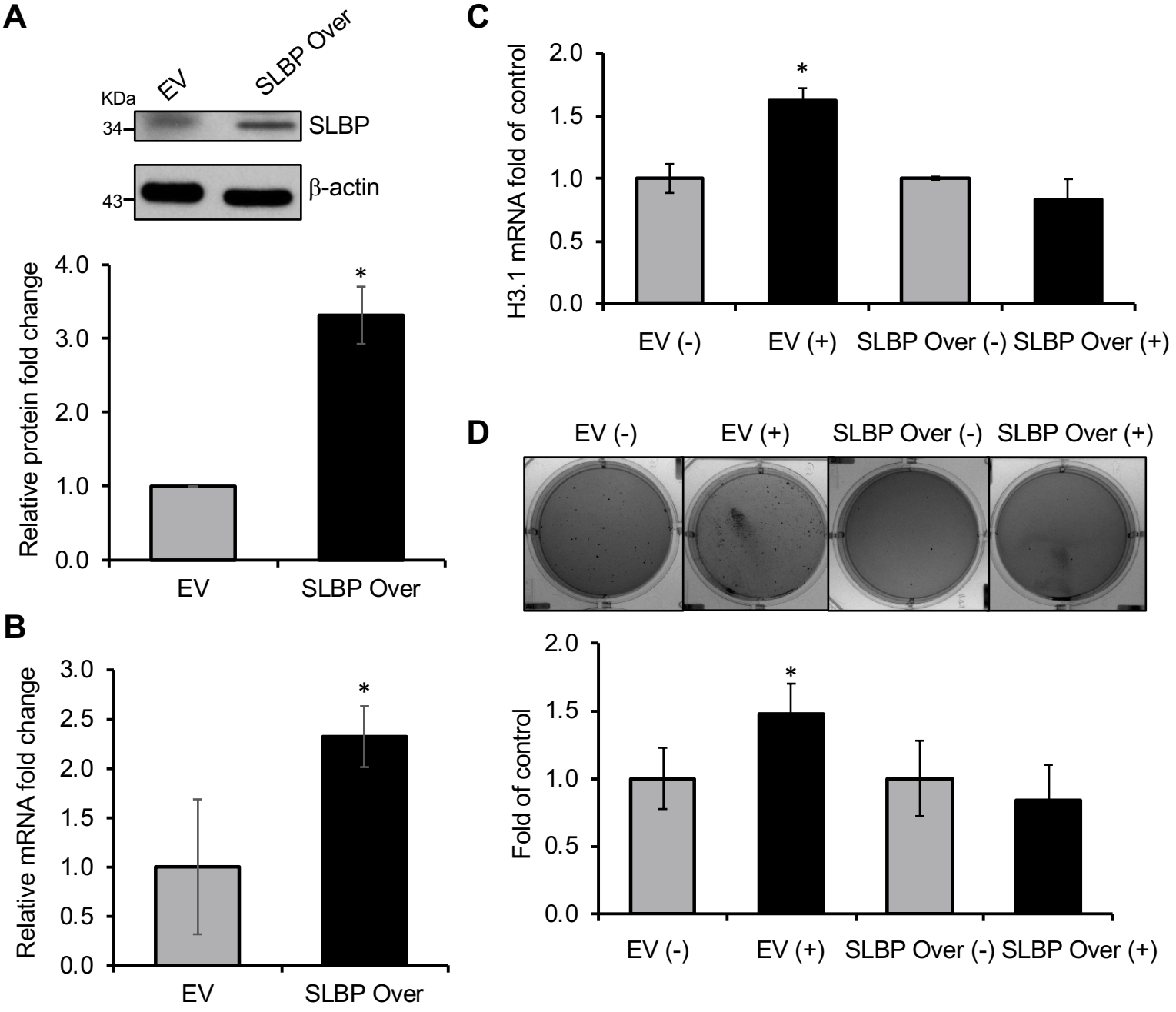

Figure S6, related to Figure 6. SLBP overexpression rescues arsenic-induced cell transformation

(A) BEAS-2B cells were stably transfected with pcDNA-empty (EV) or with pcDNA-SLBP (SLBP) plasmid. Western blot was performed using the indicated antibodies. The bar graph shows relative quantification of SLBP levels normalized to β-actin. The data shown are the mean ±S.D. (n=3). Student's t test was applied for statistical significance: * $p<0.05$.

(B) mRNA levels of *SLBP* were analyzed by RT-qPCR and were normalized to *β-actin*. Data are mean ±S.D. (n=3). Student's t test was applied for statistical significance: * $p<0.05$.

(C) RT-qPCR analysis of polyadenylated H3.1 mRNA level. mRNA levels for H3.1 were normalized to *β-actin*. The data shown are the mean ±S.D. (n=3). Student's t test was applied for statistical significance: * $p<0.05$.

(D) Soft agar assays. The data shown are the mean ±S.D. (n=3). Student's t test was applied for statistical significance: * $p<0.05$.

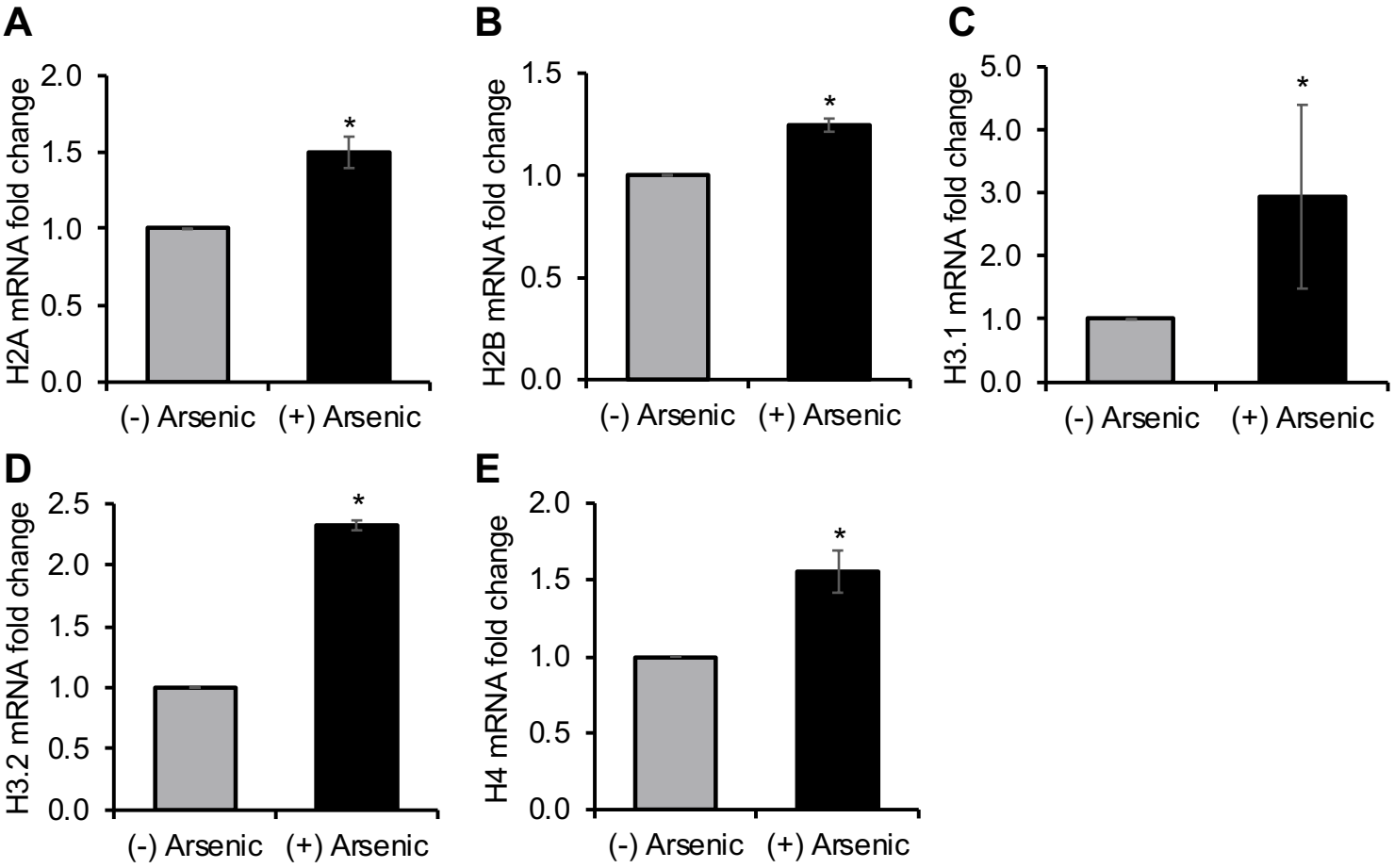

Figure S7, related to Figure 7. Arsenic induces mRNA polyadenylation of all canonical histones

(A-E) BEAS-2B cells were treated with 1 μM arsenic for 96 hours. Relative polyadenylated mRNA levels of *H2A* (A), *H2B* (B), *H3.1* (C), *H3.2* (D), and *H4* (E) were determined by RT-qPCR with oligo(dT) primers. Values are presented as mean ±S.D. from the experiments performed in triplicate. Student's t test was applied for statistical significance: * $p<0.05$.

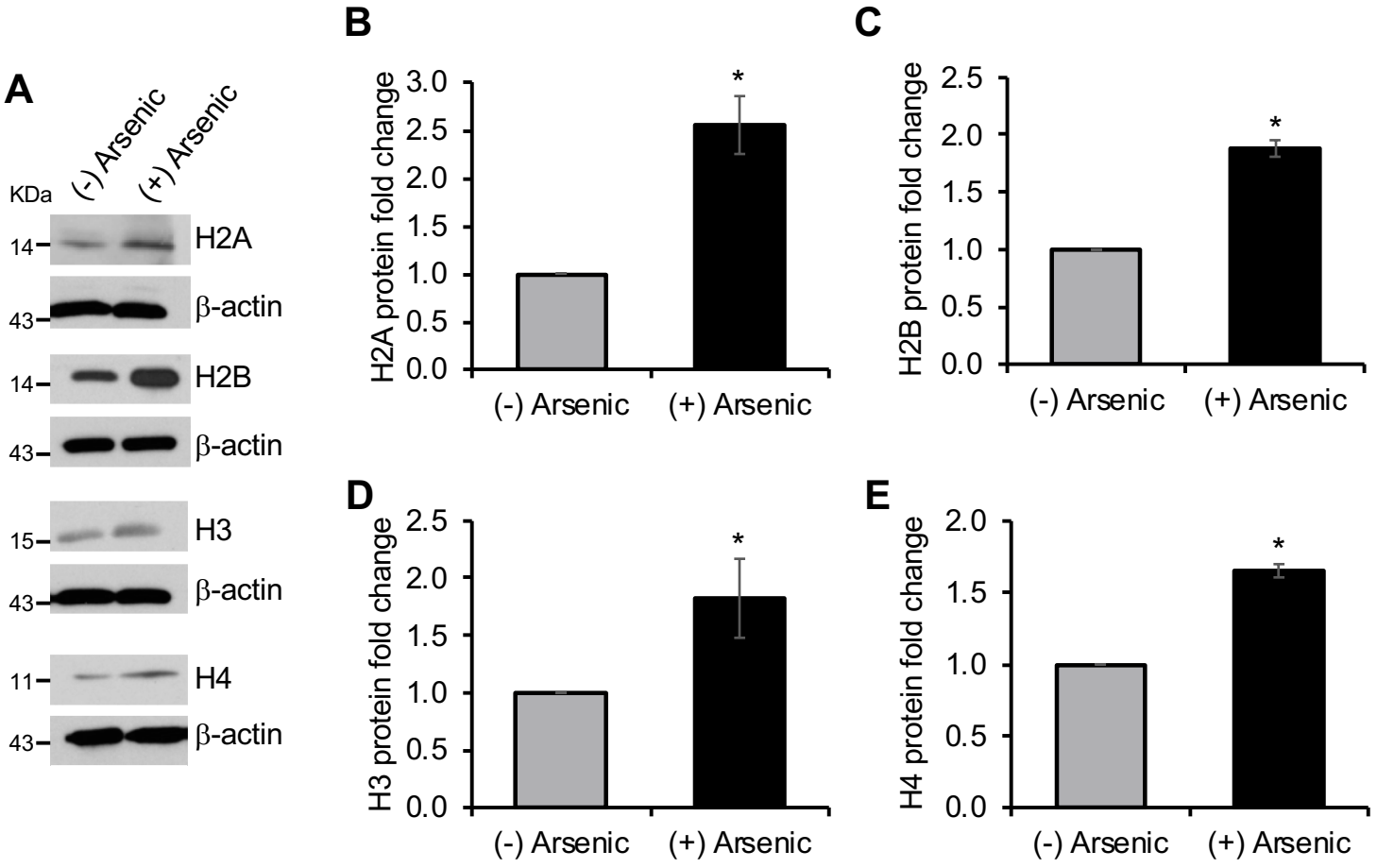

Figure S8, related to Figure 1C. Arsenic induces canonical histone protein elevation

(A) Arsenic treated BEAS-2B cells were lysed and immunoblotted against β-actin, H2A, H2B, H3, or H4, separately.

(B-E) The bar graphs show relative quantification of canonical histone levels normalized to β-actin. The data shown are the mean ±S.D. from the experiments performed in triplicate. Student's t test was applied for statistical significance: *p<0.05.

Transparent Methods

Cell Lines

Immortalized human bronchial epithelial (BEAS-2B) cells were obtained from the American Type Culture Collection (ATCC, Manassas, VA). BEAS-2B cells were adapted to serum growth immediately after purchase and have been carefully maintained at below confluent density. The cells were recently tested by Genetica DNA Laboratories and found to be 100% authentic against a reference BEAS-2B cell line (Burlington, NC). Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen, Grand Island, NY) supplemented with 1% penicillin/streptomycin (GIBCO, Grand Island, NY) and 10% heat-inactivated fetal bovine serum (FBS, Atlanta Biologicals, Lawrenceville, GA). All cells were cultured in a 37°C incubator containing 5% CO_2. BEAS-2B cells were authenticated by Genetica DNA Laboratories (Burlington, NC) on July 22, 2015. Cells were matched 100% to 15 short tandem repeat (STR) loci and amelogenin to the reference profile of BEAS-2B (ATCC CRL-9609). For arsenic exposure, cells were treated with sodium meta-arsenite ($NaAsO_2$, Sigma, St. Louis, MO) with doses ranging from 0 to 2 μM for 0 to 96 hr.

Human papillomavirus transformed prostate epithelial cell line PZ-HPV-7 was obtained from the ATCC. Cells were cultured in Keratinocyte Serum Free Medium (K-SFM) (GIBCO, Grand Island, NY), supplemented with 0.05 mg/ml bovine pituitary extract (BPE) and 5 ng/ml human recombinant epidermal growth factor (EGF), as well as 1% penicillin/streptomycin solution (GIBCO, Grand Island, NY), in a 37°C incubator containing 5% CO_2. Human embryonic kidney cell line HEK293 was obtained from the ATCC. Cells were cultured in DMEM (Invitrogen, Grand Island, NY) supplemented with 1% penicillin/streptomycin (GIBCO, Grand Island, NY) and 10% fetal bovine serum (Atlanta Biologicals, Lawrenceville, GA) in a 37°C incubator containing 5% CO_2.

Animals, arsenic exposure, and sample collection

7-week-old male A/J mice (SPF grade) were purchased from Jackson Laboratories (Bar Harbor, ME) and maintained at the New York University School of Medicine animal facility under standard environmental conditions (22 °C, 40-70% humidity, and a 12:12-hour light: dark cycle). All mice were handled in accordance with NIH and Institutional Animal Care and Use Committee (IACUC) guidelines. A purified rodent diet was provided. The mice were observed for one week before the start of the experiment. Sodium arsenite solutions were prepared at 0, 100, and 200 μg/L concentrations, and were given to mice (n=8 for each dose) via oropharyngeal aspiration every other day for 1 week. In brief, 100 μl of a solution was placed at the back of the tongue of an anesthetized mouse for aspiration into the lung in a Biosafety Level II containment facility. Aspiration is considered the best way to get uniform distribution of the soluble carcinogens to all parts of the lung. Post-treatment, the mice were sacrificed by CO_2 and lung tissues were extracted. Protein lysates were extracted using boiling buffer and analyzed for SLBP levels using Western blotting using β-actin as internal control. RNA was extracted using standard Trizol protocol and

analyzed for H3.1 poly(A) and *SLBP* mRNA using qPCR using *GAPDH* as loading control. All procedures were conducted in compliance with New York University's guidelines for ethical animal research and the Declaration of Helsinki.

Xenograft tumor model

18 Athymic NCr-nu/nu (5-6-week-old) mice were obtained from the National Cancer Institute (NCI, Frederick, MD) and housed in a pathogen-free facility for all experiments. A total of 5 million control or H3.1poly(A)-transfected cells were suspended in 0.1 mL PBS and subcutaneously injected into each side of the femoral area of the same mouse. Each mouse was injected on one side with control and the other with H3.1-transformed cells to ensure the same biological environment. Each cell line was subcutaneously injected into three mice, in triplicates. Tumor growth and overall health of the mice were monitored once a week. At 5 months post-injection, the mice were sacrificed by CO_2 euthanization and the tumors were extracted by standard surgery for determination of tumor weight. Cut-off weight for tumor weight was 0.1g. Tumor size was measured using calipers at the indicated times. The tumor volume was calculated according to the formula: volume = length*(width^2)/2. Graphs illustrating tumor number, weight, and volume were generated using Prism 7 7.0e (GraphPad Software). Animal experiments in the present study were performed in compliance with the guidelines of The Center for Laboratory Animal Research, New York University School of Medicine.

Plasmids and Cell Transfection

pcDNA3.1-H3.1-poly(A) and pcDNA3.1-FLAG-H3.1-poly(A) plasmids were previously constructed in our lab. A 110 bp DNA fragment of H3.1 gene containing the stem loop sequence and 5′ and 3′ linker sequences for sub-cloning into the XbaI and ApaI sites were amplified by PCR using the human genomic DNA as a template. The fragment was inserted into the pcDNA3.1-H3.1-poly(A) plasmid to obtain pcDNA3.1-H3.1-Loop plasmids. Transfections were carried out using PolyJet (SignaGen, Rockville, MD) according to the manufacturer's instructions. pcDNA3.1-empty vector was used as the control. Primers: forward 5′-CTAGTCTAGAGTCTGCCCGTTTCTTCCTC-3′; reverse5′-CCGGGCCCAAACTAACATAGACAACCGA-3′.

Control short interfering RNA (siRNA) or two distinct siRNAs for H3.3 (H3.3 siRNA1 and H3.3 siRNA 2) (Sigma, St. Louis, MO) were transfected into BEAS-2B cells to knock down H3.3 expression. Transfections were carried out using LipoJet (SignaGen, Rockville, MD) according to the manufacturer's instructions.

pcDNA-FLAG-SLBP plasmids were purified using a Qiagen QIAprep Spin Midiprep kit prior to transfection. Overexpression transfections were performed using Lipofectamine® LTX Reagent with PLUS reagent (Invitrogen, Grand Island, NY) following the manufacturer's protocol. Briefly, 150,000 cells were seeded into 6-well dishes 24 hours prior to transfection. The following day, 1 μg of purified plasmid was transfected into each well using 10 μL of Lipofectamine LTX and 2.5 μL of PLUS reagent per transfection. 24 hours post-transfection, the media was removed and

replaced with fresh DMEM. After three days, 0.5 μg/ml of G418 selection agent was added to the transfected cells. The cells were grown under selection for three weeks and harvested for western blot and qPCR analysis.

Antibodies

Name	Source	Catalog Number
Histone H3	Abcam	ab1791
SLBP	Abcam	ab 181972
β-actin	Abcam	ab8226
Histone H3.3	Invitrogen	MA5-24667
Histone H4	Millipore	07-108
Histone H2A	Millipore	07-146
Histone H2B	Millipore	07-371
FLAG	Sigma	F3165

Western blotting

Cells were lysed with boiling buffer (1% SDS, 10 mM Tris (pH 7.4), 1 mM sodium orthovanadate) and 50 μg of whole cell lysate were separated by 12% SDS-PAGE and transferred to a PVDF membrane. After blocking in 5% skim milk in TBST for 1 h at room temperature, the membrane was incubated with primary antibodies overnight at 4 °C, and then probed with HRP labeled secondary antibody (1:2000) for 1 h at room temperature before the visualization by the chemiluminescence. Quantification of immunodetected proteins was performed using Image J software.

RNA Isolation and Quantitative Real-time PCR

Total RNA was extracted using TRIzol (Invitrogen, Grand Island, NY) and subsequently synthesized into single-stranded cDNA using ProtoScript® II First Strand cDNA Synthesis Kit (New England BioLabs, Ipswich, MA) in accordance with manufacturer's instructions. 1μg of RNA in a final volume of 20 μl. Quantitative real time PCR analysis was performed using Power SYBR Green PCR Master Mix (Qiagen) on the ABI PRISM 7900HT system. All experiments were performed in triplicates. Relative gene expression levels were normalized to either Tubulin or GAPDH expression. The results were presented as fold change to the level expressed in control cells.

Primers Used

Gene ID	Primer Sequence	
H3.1c	Forward	5′-AAATCGCCCAGGACTTCAAA-3′
	Reverse	5′-TCCTGCAGCGCCATCAC-3′
H2Aa1	Forward	5′-CACCATAAAGCCCAAAGCAAGTA-3′
	Reverse	5′-GACGGTAGGTGGCTCTGAAAAG-3′
H2Bd1	Forward	5′-CTCAGAAGAAGGACGGGAAGAA-3′

	Reverse	5′-CGGGATGGACCTGCTTCA-3′
H3.2c	Forward	5′-GTCCACGGAGCTGCTGAT
	Reverse	5′-GCAGGTCCGTCTTAAAGTC-3′
H4D	Forward	5′-TCCGCGATGCTGTCACCTA
	Reverse	5′-TCCATGGCTGTGACTGTCTTG-3′
SLBP	Forward	5′-CAGTCTTGCCACAACTTCAATC-3′
	Reverse	5′-ATGGAGCCGATTATGAGAACAC-3′
Tubulin	Forward	5′-CGGCTGAATGACAGGTATCCTAAG -3′
	Reverse	5′-CTCGTCCTGGTTGGGAAACA -3′
GAPDH	Forward	5′-TGCACCACCAACTGCTTAGC-3′
	Reverse	5′-GGCATGGACTGTGGTCATGA-3′

RNA stability assay

Actinomycin D can inhibit transcription *in vivo* and is a simple method to measure mRNA kinetics. $1x10^5$ cells were counted for each well of a 6-well plate. After seeding, cells were treated with Actinomycin D (Invitrogen, Grand Island, NY) at 10 μg/ml for the indicated time (0, 0.5, 1, 2, 3, 5 h). The cells were then washed by PBS twice and collected by TRI-reagent and subjected to RNA extraction and RT-qPCR analysis of exogenous H3.1 mRNA expression. The primers designed for detection of exogenous H3.1 on the vector are as follows: Forward: 5′-TCCAGTGTGGTGGAATTCTG-3′; Reverse: 5′-GCTCTTACGGGCTGCTTTA-3′.

DNA extraction and ddPCR

DNeasy Blood & Tissue (Qiagen, Austin TX) kit was used to extract DNA from cells. $5x10^6$ cells of each cell type (BEAS-2B cells stably overexpressing H3.1poly(A), H3.1Loop, and the empty vector control) were harvested and centrifuged for 5 min at 300 g, and resuspended in 200 μL PBS. 20 μL proteinase K and 200 μL of Buffer AL were added in and samples were incubated for 10 min at 50°C. 200 μL ethanol (100%) was added to the samples and passed through a DNeasy mini spin column for 1 min at 6000 g. The columns containing sample DNA were washed with AW1 and AW1 buffer, respectively, and the final product was eluted using AE buffer.

Copy number of exogenous H3.1 in the extracted DNA of BEAS-2B cells stably overexpressing histone H3.1poly(A) and H3.1Loop was detected using QX200 Droplet Digital PCR System (Bio-Rad) in the Genome Technology Center Core at NYU Langone Health. Primer-probes for exogenous H3.1 (FAM-tagged) were designed spanning the H3.1 start codon site on the vector (sequence shown below); for ribonuclease P/MRP 30kDa subunit (RPP30, Hex-tagged) were obtained from Bio-Rad (validated; dHSACP1000485). DNA was digested using HindIII according to the manufacturer's instructions (NEB R0104S). For each reaction, 10 ng of restriction digested DNA was mixed with 1X ddPCR Supermix for probes (Bio-Rad) and primer probes for both H3.1 and RPP30 to a final volume of 20 μL, as recommended by Bio-Rad. Reactions without any DNA template were also run as a control on every PCR plate. Following emulsion generation on the QX200 Automated Droplet Generator (Bio-Rad), the samples in a 96-well PCR plate were

heat-sealed with foil, and amplified in a thermal cycler for 40 cycles with an annealing temperature of 58 deg Celsius. Post PCR, the droplets were read using QX200 Droplet reader (Bio-Rad) and CNVs were determined using the QuantaSoft™ Software version 1.7 (Bio-Rad). The primers designed for exogenous H3.1 on the vector are as follows: Forward-ACTAGTCCAGTGTGGTGGAA; Reverse-GGTAGACTTGCGAGCTGTTT; and Internal probe: TATCCAGCACAGTGGCGGCC.

Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

The total RNA of treated cells was extracted using the Trizol reagent (Invitrogen, Grand Island, NY); 1 μg of RNA was reverse transcribed using the SuperScript IV First-Strand Synthesis System (Invitrogen, Grand Island, NY) following the manufacturer's protocol. Amplification cycles were: 95 °C for 5 min and then 30 cycles at 95 °C for 30 s, 58 °C for 30 s, 72 °C for 30 s, followed by 72 °C for 10 min. Aliquots of PCR products were checked by electrophoresis on a 2% agarose gel and the fragments were visualized by ethidium bromide staining.

Primers Used

Gene ID	Primer Sequence	
H3.1 c	Forward	5′-CGAAATCCGTCGCTACCAGA-3′
	Reverse	5′-GCGCACAGATTGGTGTCTTC-3′
T7	Forward	5′-TAATACGACTCACTATAGGG-3′
BGH	Reverse	5′-TAGAAGGCACAGTCGAGG-3′

Cell proliferation assay

Cells were plated in 96-well plates at a density of $5x10^3$ cells per well and incubated at 37°C in 5% CO_2. At different time points (0, 1, 2, 3, 4, 5, 6, 7 days), the medium was replaced with 100 μl fresh medium containing 0.5 mg/ml 3-(4,5-dimethylthiazol-2-ul)-2,5-diphenyl tetrasodium bromide (MTT). Four hours after the addition of MTT, 100 μl of isopropanol containing HCL was added to each well to dissolve the crystals. Fluorescence was monitored using a microplate reader (Spectra Max M2, Molecular Devices) at a wavelength of 570 nm. Each experiment was performed eight times and repeated three times. Data was reported as mean ± SD.

Anchorage independent growth assay

Cells were rinsed with PBS to remove the metal from the media then seeded in low gelling temperature Agarose Type VII (Sigma Aldrich, St. Louis, MO). 5,000 cells were seeded in triplicate in 6-well plates in a top layer of 0.35% agarose onto a bottom layer of 0.5% agarose. BEAS-2B cells and PZ-HPV-7 were allowed to grow for four weeks until individual colonies were large enough to be selected from the agar. HEK293 cells were allowed to grow for two weeks. Colonies were picked from each treatment and control group. These colonies were grown out into monolayers for four weeks. After monolayer growth, cells were collected in Trizol for quantitative real-time PCR (RT-qPCR). A second set of plates was stained with INT/BCIP solution (Roche

Diagnostics, Indianapolis, IN) overnight for visualization and quantification of colony growth in agar, according to the manufacturer's protocol.

Flow cytometry

For flow cytometry cell cycle analysis, the cells were fixed with 4% formaldehyde for 10 min at RT and permeabilized by adding cell suspension drop-wise into ice-cold 100% methanol with gentle vortexing to a final concentration of 90% methanol and incubated on ice for 30 min. Permeabilized cells were resuspended in FACS buffer (1xPBS, 1% BSA), stained with Alexa-488-conjugated phospho-Histone H3 (ser10) (1:50, CST9708, Cell Signaling Technology, Danvers, MA) for 1 hr at room temperature in the dark. Then cells were counterstained with propidium iodide (Sigma-Aldrich, St. Louis, MO) for 30 min at room temperature in the dark. Stained cells were counted and analyzed via FACS Calibur flow cytometer (BD, Bioscience, San Jose, CA) and FlowJo (Ashland, OR).

Transwell cell invasion assay

For cell invasion analysis, Biocoat Matrigel® Invasion Chambers (Corning, Corning, NY) were incubated with DMEM containing 0.1% FBS at 37°C for 3 h. BEAS-2B cells were trypsinized, centrifuged, and washed to remove excess FBS. $5x10^4$ cells were resuspended with 400 μL of DMEM containing 0.1% FBS. The cells containing DMEM with 0.1% FBS were seeded in wells containing DMEM with 10% FBS and grown for 24 h in a 37°C incubator containing 5% CO_2. The migrated cells were fixed with 3.5% formaldehyde for 5 min at room temperature and subsequently incubated with methanol for 20 min. The wells were then incubated with 5% Giemsa solution overnight at room temperature and washed with water the next day for observation under the microscope.

Chromosome spread analysis

Transfected cells were arrested in metaphase by the addition of 0.1 μg/ml colcemid for 3 hr. The cells were trypsinized and collected by centrifugation at 1,000 rpm 5 min. Hypotonic treatment was accomplished with 0.075 M KCl at 37°C for 17 min, followed by fixation in 3:1 methanol–acetic acid (three changes). Prepared cells were dropped on cold, wet slides, stained for 5 min in 10% Giemsa (Sigma-Aldrich, St. Louis, MO) in pH 7.0 phosphate buffer and observed under the light microscope with oil lens. Chromosomes spreads were scored blinded according to the following criteria: deletions, breaks, dicentric, ring, trisome and etc. The results were represented as aberrations per 60 chromosome spreads.

Nucleosome preparation, ChIP, and ChIP-seq

Mono- and dinucleosomes were isolated by Micrococcal Nuclease (MNase) digestion and sucrose gradient purification as described (Jin and Felsenfeld, 2006). ChIP was performed using anti-FLAG M2 Affinity-gel (A2220, Sigma-Aldrich, St. Louis, MO) or anti-H3.3 antibody (MA5-24667, Invitrogen) (Jin and Felsenfeld, 2007). ChIP-seq libraries were prepared using an Illumina

TruSeq ChIP sample preparation kit (IP-202-1024, Illumina, San Diego, CA) according to the manufacturer's protocol. Sequencing was performed either with the IIlumina HiSeq 2500 platform to obtain 50-nucleotide single end reads for FLAG ChIP at the NYULMC Genome Technology Center or with the Illumina HiSeq 4000 platform to get 2x150 bp paired-end reads for H3.3 ChIP by GeneWiz Inc. After sequencing, the Illumina reads were mapped to the human genome. Regions of enrichment were identified using MACS peak calling algorithm and GREAT database was used to annotate the aberrant enriched regions.

RNA sequencing

Total RNA from stable untagged H3.1poly(A) cells was converted into complementary DNA (cDNA) libraries using a Truseq RNA Sample Preparation V2 Kit (Illumina, San Diego, CA). Validation of library preparations was performed on an Agilent Bioanalyzer using the DNA1000 kit. Library concentrations were adjusted to 4 nM, and libraries were pooled for multiplex sequencing. Pooled libraries were denatured and diluted to 15 pM and then clonally clustered onto the sequencing flow cell using the Illumina cBOT Cluster Generation Station and a TruSeq Paired-End Cluster Kit v3-cBot-HS. Sequencing was performed on an Illumina HiSeq2500 Sequencing System using a TruSeq SBS Kit v3-HS.

Statistical analysis

Image J processing software (National Institutes of Health) was used to quantify western blotting gel intensities. All statistical significance was calculated and assessed using unpaired, 2 tailed t-test, where * indicates $p<0.05$ and ** indicates $p<0.01$.

DATA PROCESSING

ChIP-seq data analysis

Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to trim the raw sequence reads ($ trim_galore --phred33 --fastqc --clip_R1 5 --three_prime_clip_R1 3 R1.fq -o OUTDIR). Reads were aligned to the human reference genome (GRCH38/hg38) using BWA (Li and Durbin, 2009) ($bwa mem -t 8 INDEX IN.fq > PRENAME.sam). SAM files were then converted into BAM format using samtools (Li et al., 2009) ($ samtools view -bS -q 1 -@ 8 PRENAME.sam > PRENAME.bam). Bedtools was used to convert BAM files into BED format ($bedtools bamtobed -i PRENAME.bam > PRENAME.bed). Only 1 copy of the redundant reads that were mapped to the exact same location in the genome was retained. All non-redundant reads from FLAG-tagged or untagged (FH3.3, FH3.3NT3.1, NTH3.1 etc.). ChIP-seq experiments were included for downstream analyses without peak calling. MACS2 (Zhang et al., 2008) was used to call peaks under the FDR threshold of 0.01 ($ macs2 callpeak --SPMR -B -q 0.01 --keep-dup 1 -g hs -t PRENAME.bam -n PRENAME --outidr OUTDIR) to identify putative insulators from CTCF ChIP-seq data.

RNA-seq data analysis

RNA-seq data were processed using Salmon (Patro et al., 2017) ($ salmon quant --gcBias -i INDEX -l A -p 8 -r IN.fq -o OUTDIR). Transcriptome index was built on the human reference genome (GRCH38/hg38). Transcript-level abundance estimates were summarized to the gene-level using the R package tximport(Soneson et al., 2015) for differential expression analysis. Differentially expressed genes (DEGs) were identified using DESeq2 (Love et al., 2014) by comparing transcriptomes in H3.1poly(A)-transfected cells and in empty vector-transfected cells as well as in H3.1Loop-transfected cells and in empty vector-transfected cells, with fold-change cutoffs at 1.2 ($\log_2$FC > 0.263 or $\log_2$FC < -0.263) and adjusted P-value cutoff of 0.1. Two groups of genes were focused in the analysis: (A) genes that are up-regulated in H3.1poly(A) but are either not differentially expressed in H3.1Loop or up-regulated in H3.1Loop with a low fold change (<0.9 in H3.1poly(A)), and (B) genes that are down-regulated in H3.1poly(A) but are either not differentially expressed in H3.1Loop or down-regulated in H3.1Loop with a low fold change (<0.9 in H3.1poly(A)). 380 group A genes and 326 group B genes were identified respectively.

DNase-seq data analysis

DNase-seq data were processed using Chilin (Qin et al., 2016) with default parameters ($ chilin simple -p narrow [--pe] -s hg38 --threads 8 -t IN.fq -i PRENAME -o OUTDIR).

Composite plot of ChIP-seq profiles at enhancers and insulators

Putative enhancers were identified as top 10,000 shared MACS peaks in two A549 DNase-seq datasets (GSM736580, GSM736506) and are at least 2 kb away from any TSS. Insulators were identified as MACS peaks from CTCF ChIP-seq in BEAS-2B cell line (GSM1354438) and are at least 2 kb away from any TSS. Each ChIP-seq composite plot covers a region centered at the peak summits (2 kb for DNase-seq peaks, 1 kb for CTCF ChIP-seq peaks), and relative ChIP-seq signals (average RPKM in ChIP divided by average RPKM in chromatin input) per 20 bps bin were plotted accordingly.