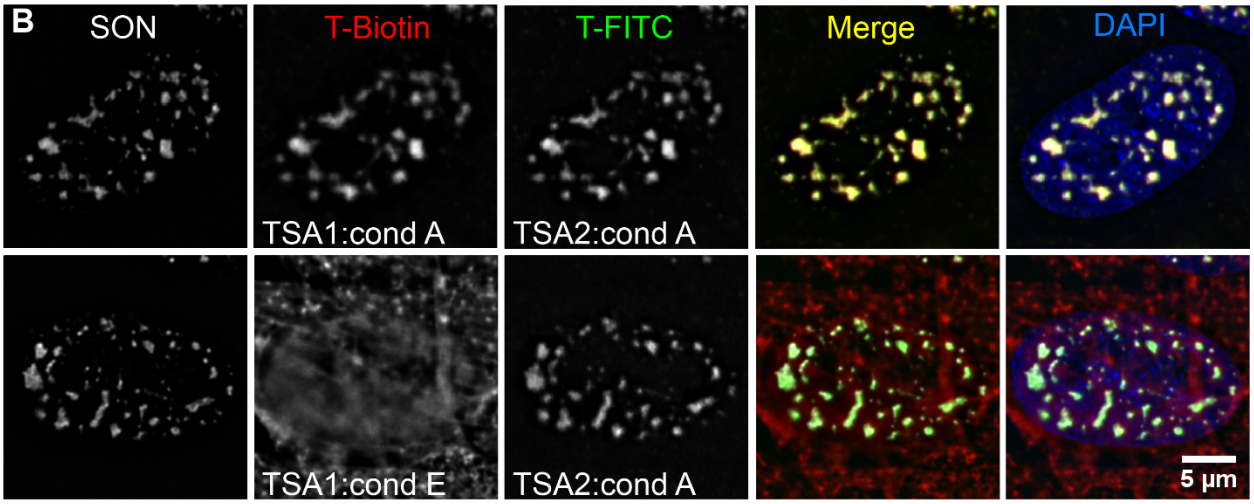
Supplemental Figures

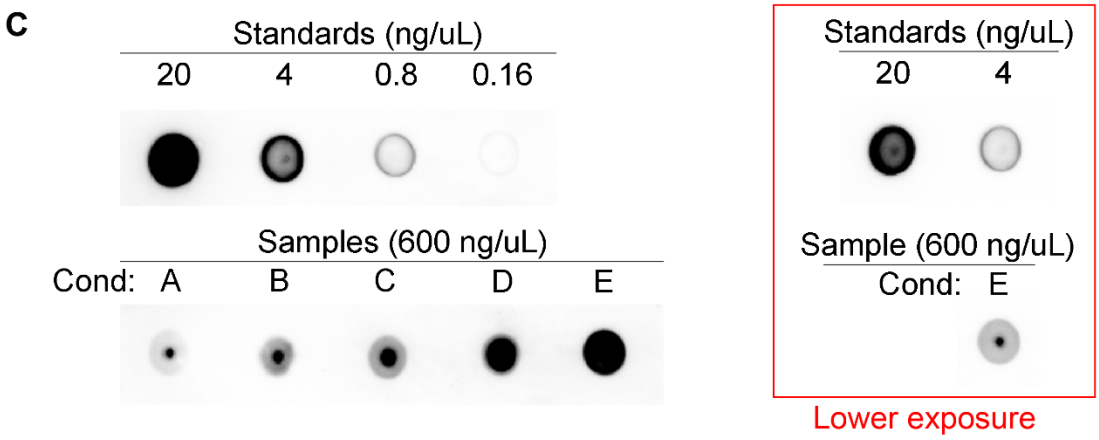
A **Condition:**

A → B → C → D → E

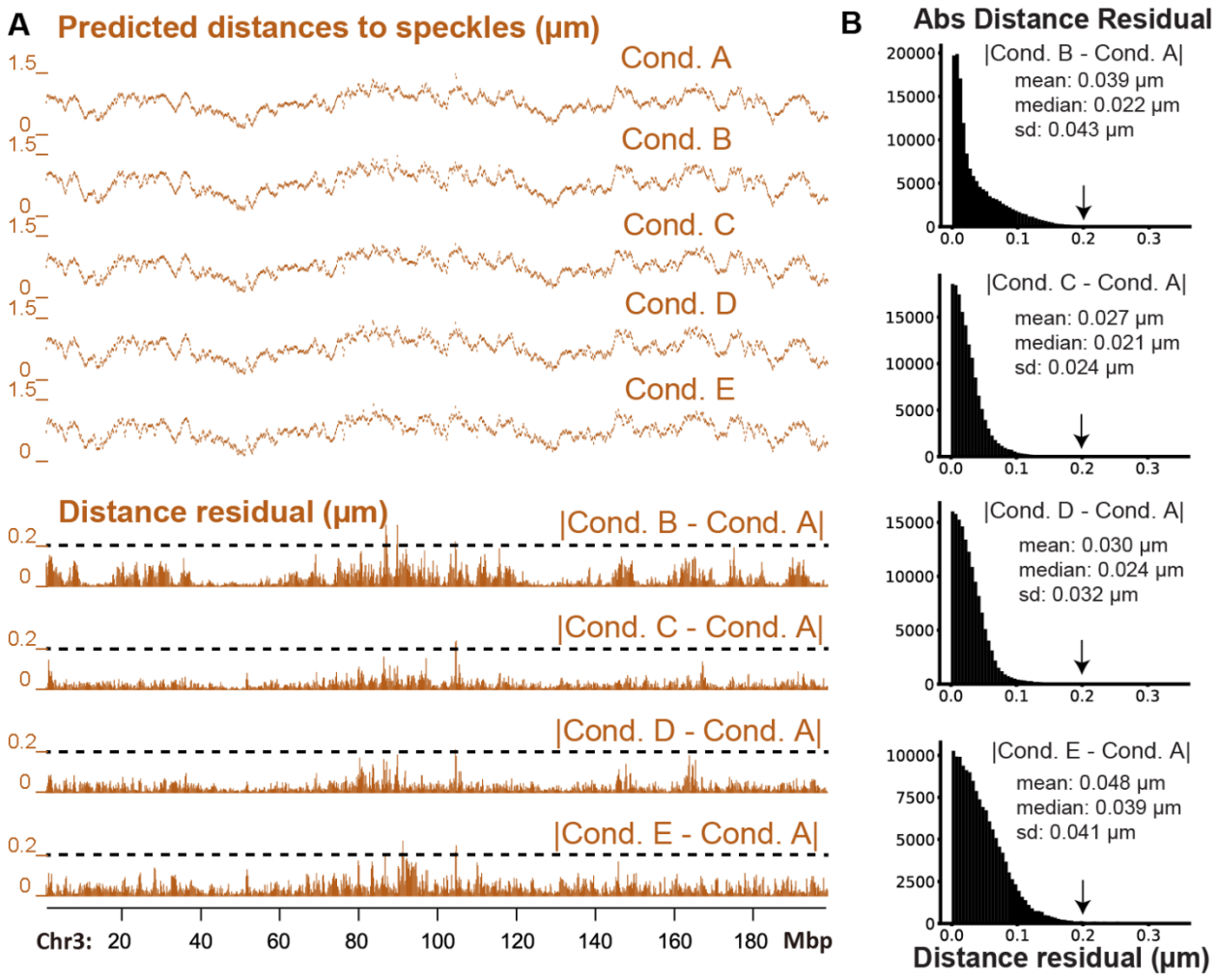
3.3X tyramide conc.	3X reaction time	3X tyramide conc.	3.3X tyramide conc.

Condition	Sucrose	Tyramide-Biotin	Hydrogen peroxide	Reaction time
A	50%	1:10000	0.0015%	10 min
B	50%	1:3000	0.0015%	10 min
C	50%	1:3000	0.0015%	30 min
D	50%	1:1000	0.0015%	30 min
E	50%	1:300	0.0015%	30 min

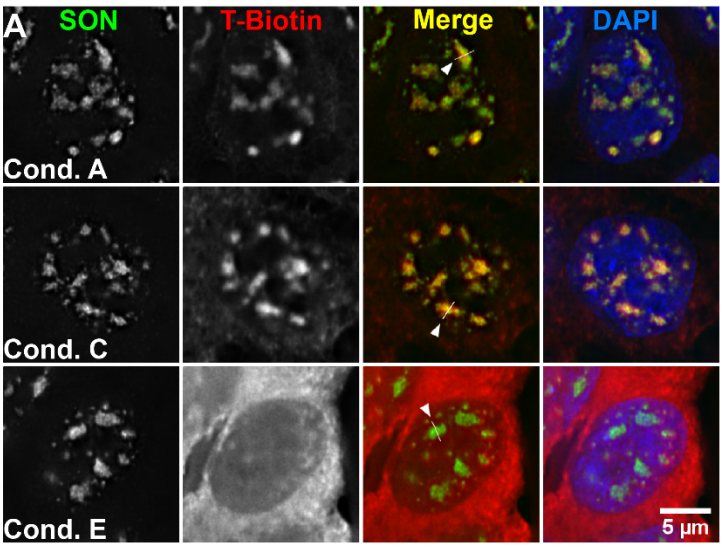


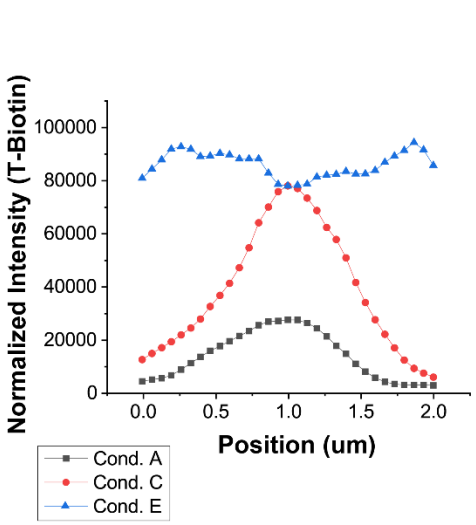


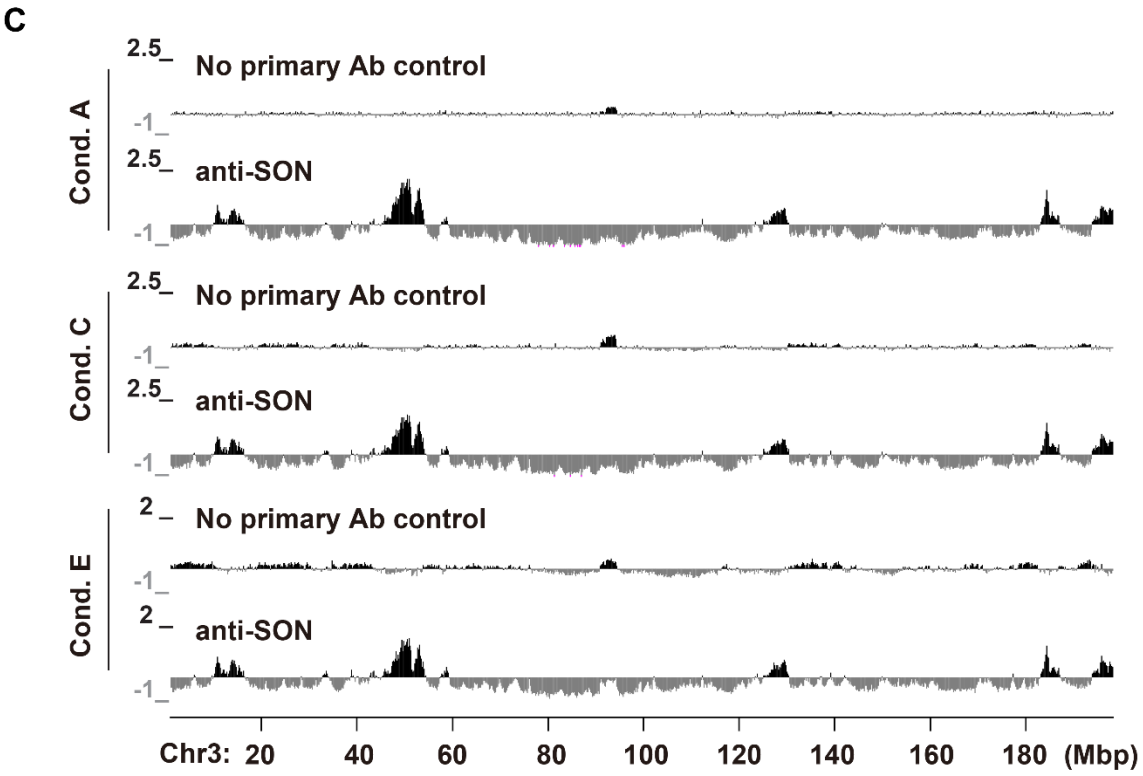
Supplemental Fig S1. TSA-seq 2.0 development and evaluation. A) Experimental design of enhanced TSA labeling conditions. All labeling conditions include 50% sucrose (w/v) and 0.0015% hydrogen peroxide (v/v). From Conditions A to E, serial increases in tyramide-biotin concentration and/or longer reaction times are applied. **B)** Two sequential rounds of SON TSA-labeling demonstrate speckle-specific DNA TSA-labeling after previous "super-saturation" TSA staining in HFFc6. Top row: Two sequential Condition A (non-saturating) rounds of SON TSA-labeling reveals speckle-specific TSA-labeling for both round 1 (tyramide-biotin) and round 2 (tyramide-FITC). Bottom row: First round of SON TSA-labeling using super-saturation Condition E (tyramide-biotin) produces non-specific staining due to saturation of protein-labeling, but second round with tyramide-FITC shows speckle-specific TSA-labeling consistent with sub-saturation DNA labeling: Left to right- SON immunostaining (grey), tyramide-biotin labeling (red), tyramide-FITC (green), merged channels, merged channels plus DAPI-staining (blue). **C)** Dot blot estimation of DNA-biotinylation levels after SON TSA labeling using Conditions A-E. Top: Biotinylated-DNA standards: dilutions of biotinylated PCR 250-bp product (0.16 – 20 ng/µl) containing 1 biotin / DNA molecule combined with unlabeled K562 genomic DNA (100-600 bp fragments) to a total DNA concentration of 600 ng/µl. Bottom: Sonicated genomic DNA (100-600 bp) from K562 cells after SON TSA-labeling using Conditions A-E (left to right). Lower exposure images for first two standards and condition E shown in box.



Supplemental Fig S2. SON TSA Conditions A-E produce quantitatively similar genome-wide estimates of mean distances to speckles. A) Chromosome 3 tracks showing: (Top) Mean distances (µm) to nuclear speckles in K562 cells estimated from SON TSA-seq produced using Conditions A-E. (Bottom) Absolute values of distance residuals (µm) between Condition A versus Conditions B-E. **B)** Histograms of absolute value distance residuals between Condition A versus Conditions B-E from all 20 kb bins across the genome. Number (y-axis), distance residual values (x-axis: intervals of 0.005 µm). Arrows mark 0.2 µm value shown as dotted horizontal lines in **A)**.



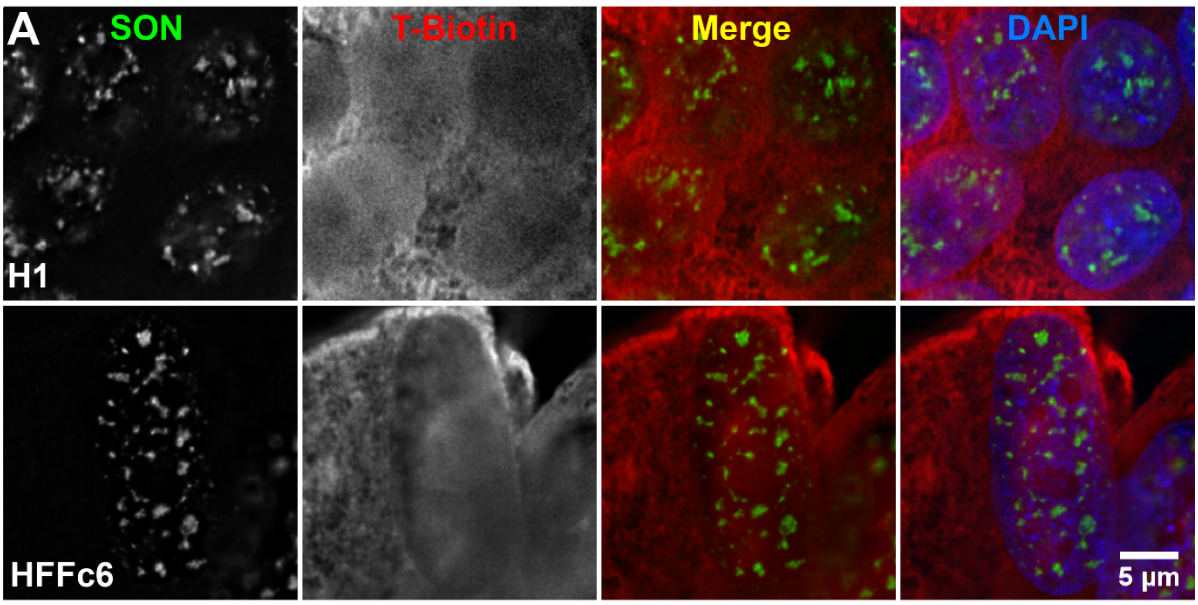


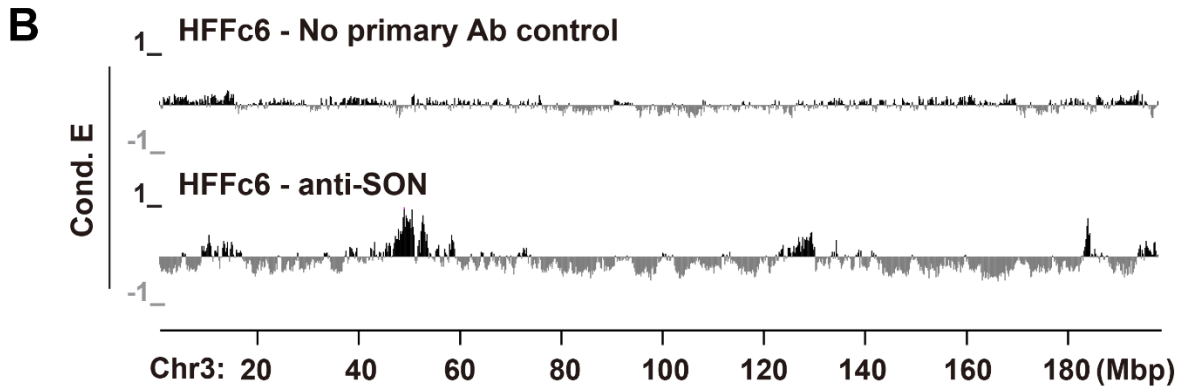


D

Cond.	Cell N (millions)	anti-SON TSA-Seq Pulldown DNA (ng)
A	440	4.8
C	220	9.3
E	220	59.3

Supplemental Fig S3. SON TSA-seq using Conditions A-E in HCT116 cells parallels results from K562 cells. A) Similar changes in reduced specificity of cellular anti-biotin staining after SON TSA in HCT116 cells with increased tyramide labeling using Conditions A (top) versus C (middle) and E (bottom). Left to right: SON immunostaining, streptavidin staining of tyramide-biotin, merged channels (SON, green; biotin, red), plus DAPI (blue). **B)** Tyramide-biotin intensities along line profiles spanning nuclear speckles in **A)** for Conditions A, C, and E. **C)** Similar SON TSA-seq enrichment score profiles (chromosome 3, 20 kb bins) for Conditions A, C, and E with no primary antibody controls. (Note the single peaks in the no primary controls map to the centromeric region and likely reflect a differential PCR-bias in this AT-rich, repetitive genomic region in the library construction of pulldown versus input DNA sequencing libraries, as several additional PCR cycles were added for the no primary control pulldowns due to the low amount of recovered DNA). **D)** Cell numbers used and pulldown DNA yields for Conditions A, C, and E.



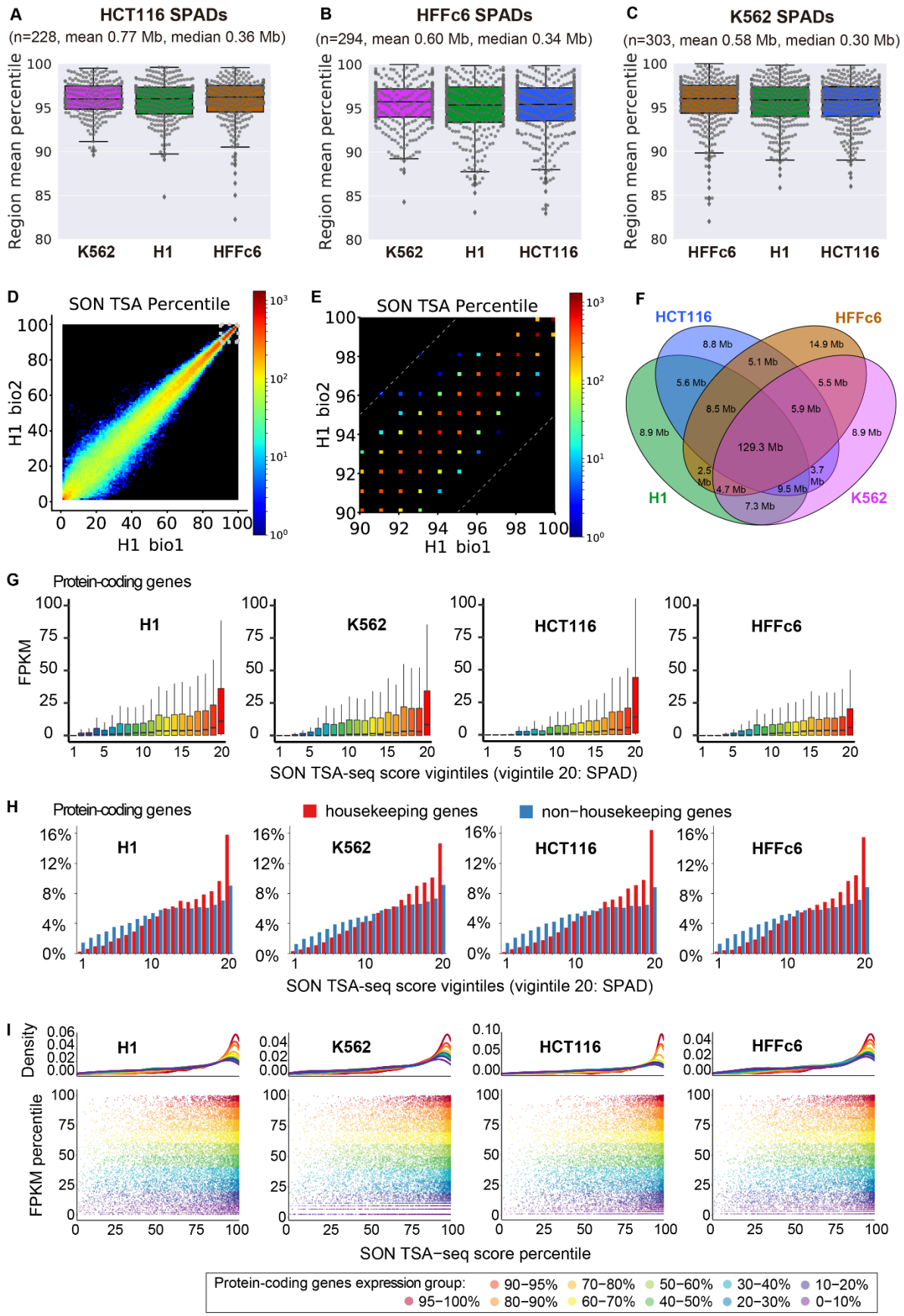


C

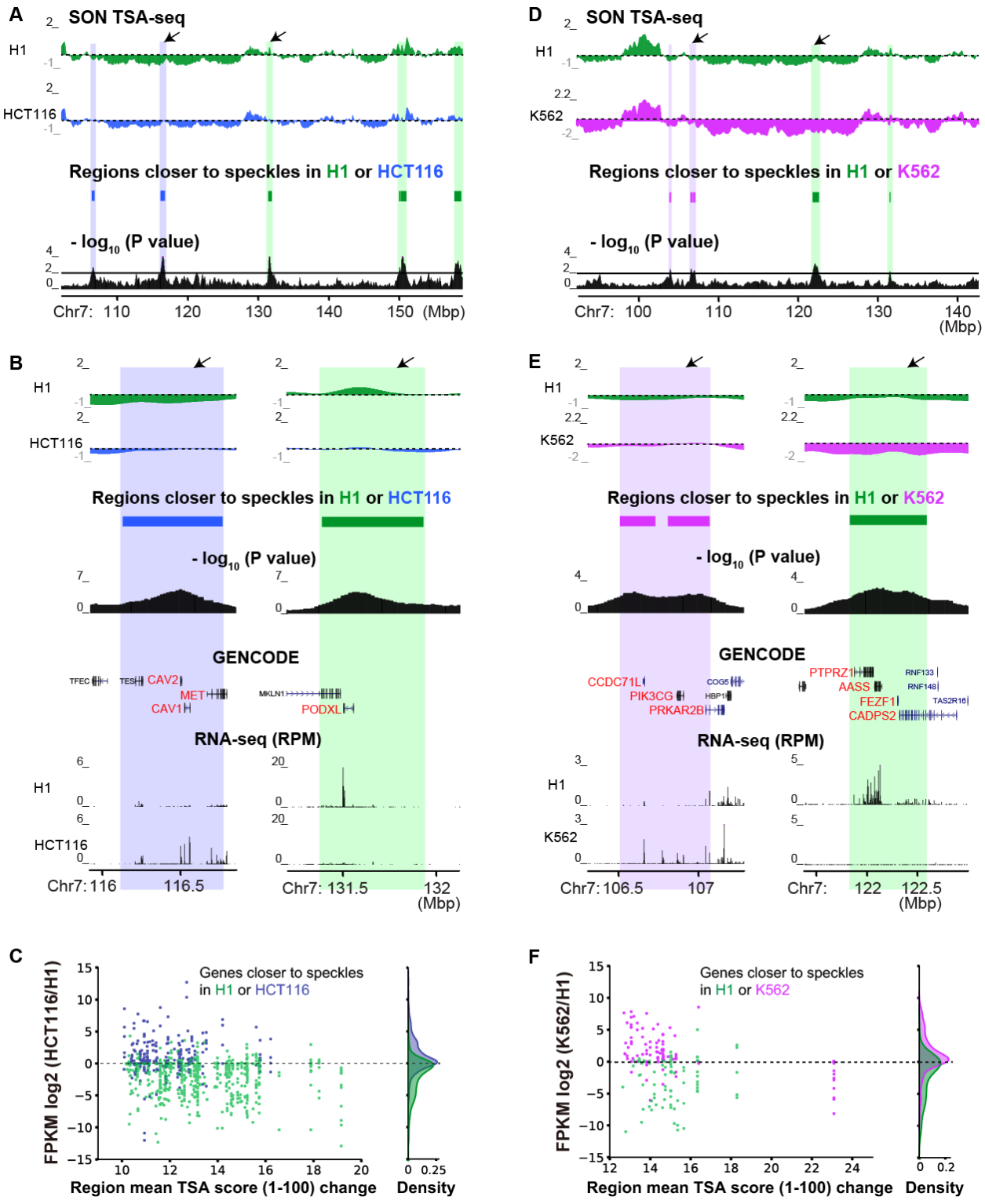
Cell line	Cond.	Cell N (millions)	Cell culture flask used	anti-SON TSA-Seq Pulldown DNA (ng)	Pulldown DNA used for NGS libary (ng)
HCT116	E	220	2 T-300	59.3	5
HCT116 (rep)	E	55	1 T-150	11.4	7.5
HFFc6	E	45	9 T-300	40.4	6
HFFc6 (rep)	E	10	2 T-300	10.8	7.5
H1	E	30	3 T-75*	60.3	7.5
H1 (rep)	E	10	1 T-75*	11.9	7.5

* We used the 4DN SOP for H1 stem cell culturing that required digesting colonies into aggregates when passaging (Methods). The estimated cell number is 10-15 million / T-75 flask at harvest.

Supplemental Fig S4. Further TSA-seq 2.0 evaluation. A) Microscopy assays of biotin-labeling after Condition E SON TSA-labeling in H1 and HFFc6. Cells were immunostained to verify biotin-labeling after Condition E SON TSA-seq: (Left to right) SON immunostaining, streptavidin staining of tyramide-biotin, merged image (SON, green; biotin, red), merged image plus DAPI (blue) in H1 (top) and HFFc6 (bottom) cells. **B)** SON TSA-seq enrichment score profiles (chromosome 3, 20 kb bins) in HFFc6 cells for Condition E with no primary antibody control. **C)** Summary of numbers of cells and culture flasks used, pulldown DNA yields, and the pulldown DNA amount used for sequencing library construction in HCT116, HFFc6 and H1 cells using TSA condition E.

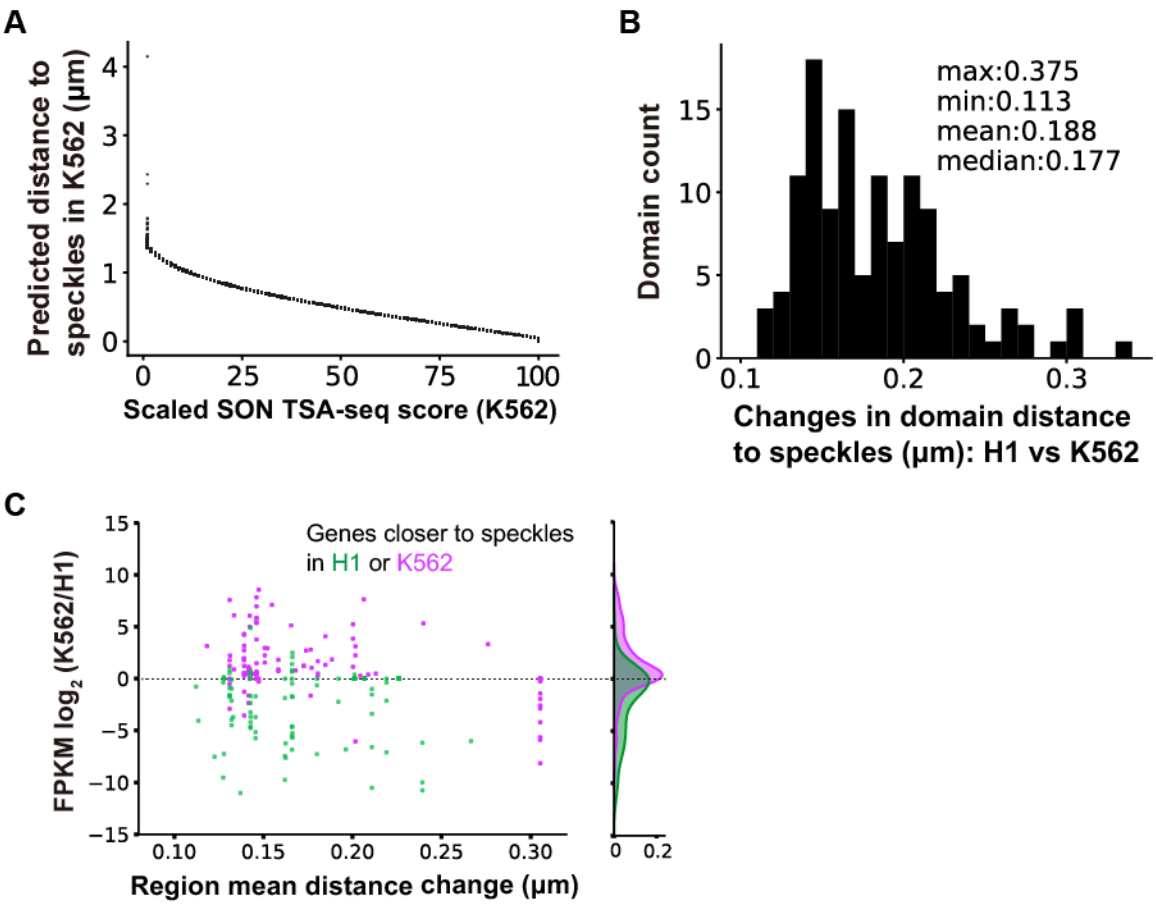

A
HCT116 SPADs
(n=228, mean 0.77 Mb, median 0.36 Mb)
Region mean percentile
K562
H1
HFFc6
B
HFFc6 SPADs
(n=294, mean 0.60 Mb, median 0.34 Mb)
K562
H1
HCT116
C
K562 SPADs
(n=303, mean 0.58 Mb, median 0.30 Mb)
HFFc6
H1
HCT116
D
SON TSA Percentile
H1 bio2
H1 bio1
E
SON TSA Percentile
H1 bio2
H1 bio1
F
HCT116
HFFc6
H1
K562
8.8 Mb
14.9 Mb
5.1 Mb
5.6 Mb
5.5 Mb
8.5 Mb
5.9 Mb
8.9 Mb
8.9 Mb
129.3 Mb
2.5 Mb
3.7 Mb
4.7 Mb
9.5 Mb
7.3 Mb
G
Protein-coding genes
FPKM
H1
K562
HCT116
HFFc6
SON TSA-seq score vigintiles (vigintile 20: SPAD)
H
Protein-coding genes
housekeeping genes
non−housekeeping genes
H1
K562
HCT116
HFFc6
SON TSA-seq score vigintiles (vigintile 20: SPAD)
I
Density
FPKM percentile
H1
K562
HCT116
HFFc6
SON TSA−seq score percentile
Protein-coding genes expression group:
90−95%
70−80%
50−60%
30−40%
10−20%
95−100%
80−90%
60−70%
40−50%
20−30%
0−10%

Supplemental Fig S5. SPADs are largely conserved and enriched for genes with higher expression levels and house-keeping genes in all four cell types. A-C) Comparison of SPADs from HCT116, HFFc6, and K562 cells with SPADs in other cell lines. TSA-seq score percentile distributions of SPADs from HCT116 (**A**), HFFc6 (**B**), and K562 (**C**) in the other 3 cell lines. Box plot displays are as described in Fig. 2B legend. **D-E)** 2D histograms show correlation of SON TSA-seq score percentiles (20 kb bins) for biological replicates 1 (x-axis) and 2 (y-axis) across entire percentile range (**D**) and top 10 percentile regions (**E**). Colors represent numbers of 20 kb genomic bins with given TSA-seq score percentiles in the 2 replicates falling within given replicate 1 percentile: replicate 2 percentile histogram 2D bin (1 percentile x 1 percentile intervals). **E)** Regions between dashed lines show histogram bins in which TSA-seq score percentiles are within a 5-percentile difference of each other in the two replicates. **F)** 4-way Venn diagram showing overlapping of SPADs across all 4 cell lines. 56.4% (129.3 Mbp) are classified as SPADs (>95th percentile) in all 4 cell lines, 12.5% (28.6 Mbp) in 3 cell lines, 13.0% in 2 cell lines (29.7 Mbp), and 18.1% (41.5 Mbp) in just 1 cell line, out of 229.1 Mbp total. However, 100% of SPADs remain near speckles (>80th percentile in relative SON TSA-seq enrichment scores) in all 4 cell lines. **G)** Protein-coding gene FPKM in 20 SON TSA-seq enrichment score vigintiles (division of percentile scores, 1-100, into 20 bins of 5% size each) in H1 (top left), K562 (top right), HCT116 (bottom left), and HFFc6 (bottom right). Box plots show median (inside line), 25th (box bottom) and 75th (box top) percentiles, 75th percentile to highest value within 1.5-fold of box height (top whisker), and 25th percentile to lowest value within 1.5-fold of box height (bottom whisker). **H)** Distributions of percentages (y-axis) of housekeeping (red) and non-housekeeping (blue) protein-coding genes versus SON TSA-seq vigintiles (x-axis) in 4 cell lines. **I)** Scatterplots for protein-coding gene FPKM percentile ranges (see color-code, y-axis) against SON TSA-seq enrichment score percentiles (x-axis) in 4 cell lines (bottom) with kernel density plots showing gene number distributions for each percentile range (top).

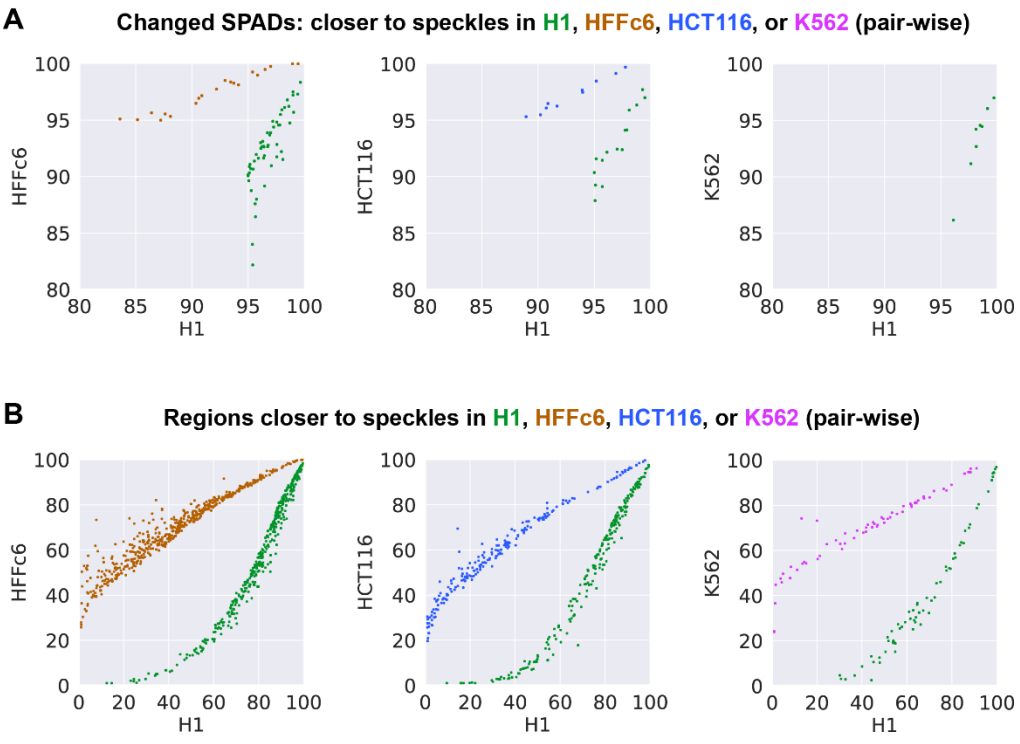


Supplemental Fig S6. Pair-wise cell type comparisons of SON TSA-seq mapping reveal regions with variable distances to nuclear speckles that are tightly correlated with changes in gene expression. A,

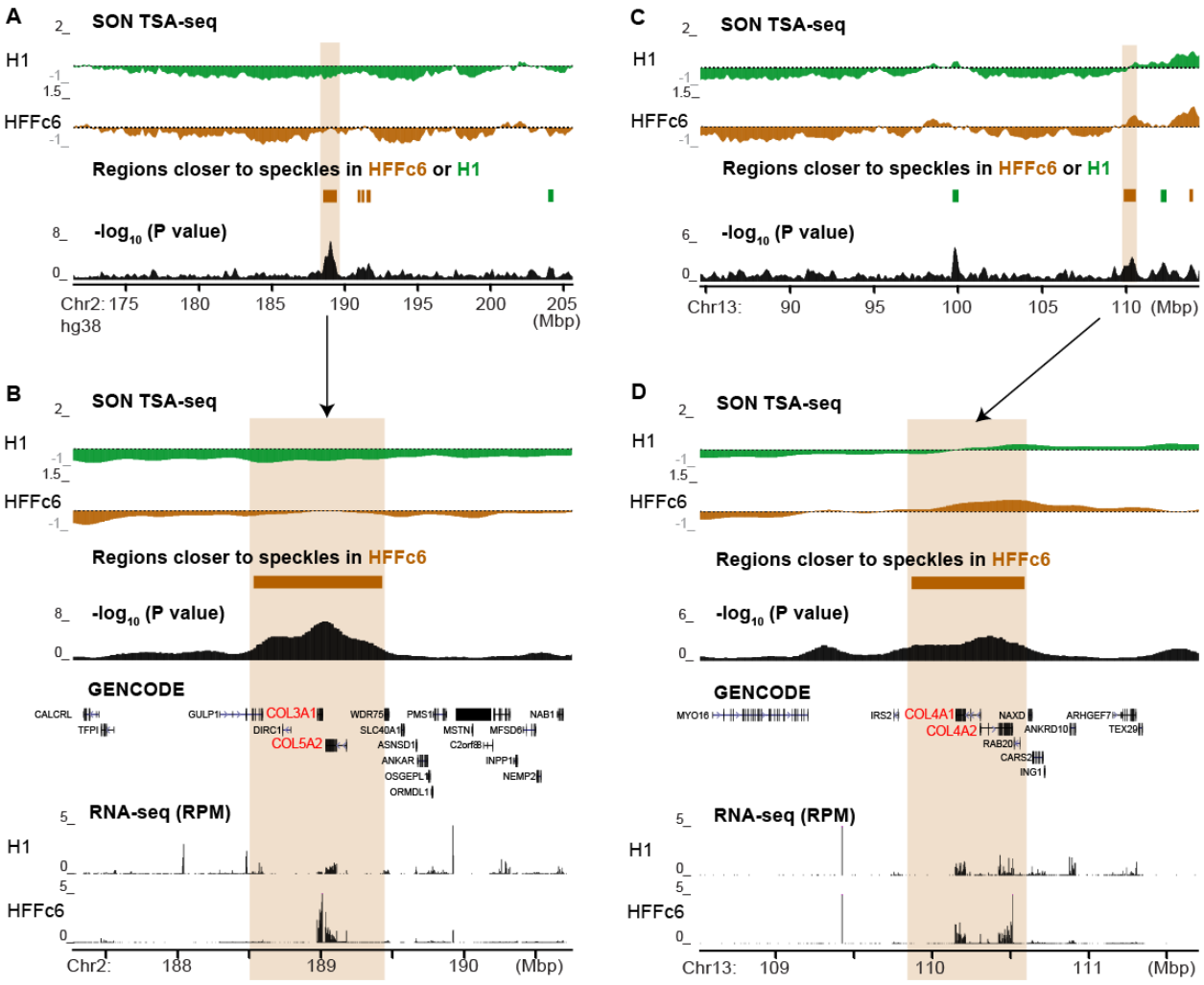
D) H1 and HCT116 (**A**) or H1 and K562 (**D**) SON TSA-seq enrichment scores (smoothed) (top, dashed lines indicate "zero" values which are regions with genome-wide average number of reads), highlighted repositioned domains (middle), and - $\log_{10}$ (p-value per bin for significance of change in scaled SON TSA-seq scores) (bottom). **B, E)** Zoomed view of two regions (arrows in **A, D**) plus gene annotation (GENCODE) and RNA-seq RPM values. **C, F)** Changes in expression of protein-coding genes in repositioned domains: scatterplots show $\log_2$ fold-changes in FPKM ratios (y-axis) between HCT116 and H1 (**C**) or between K562 and H1 (**F**) versus absolute values of changes in mean scaled TSA-seq scores (x-axis, max-min normalized: 1-100). Gene distribution is shown as Kernel density plots on the right.



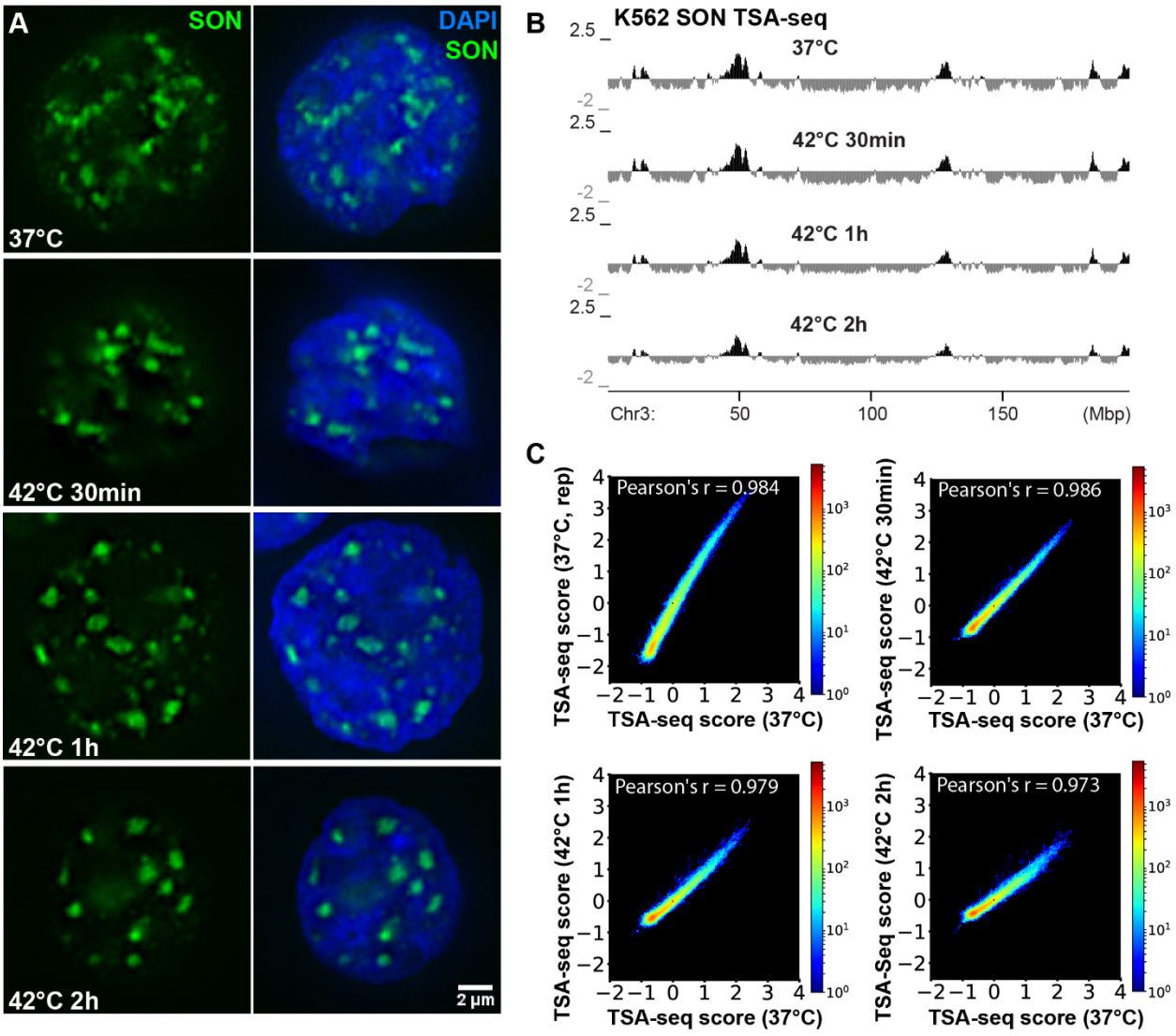
Supplemental Fig S7. Distance calibration demonstrates small distance changes relative to speckles in H1 vs K562 comparison yet a tight correlation with changes in gene expression. A) Scatter plot showing predicted distances in microns versus corresponding scaled TSA-seq enrichment scores (1-100, 20 kb bins). **B)** Distribution of predicted mean distance changes for each changed domain between H1 and K562 cells (x-axis: intervals of 0.01 µm). **C)** Changes in expression of protein-coding genes in repositioned domains: scatterplots show $\log_2$ fold-changes in FPKM ratios (y-axis) between K562 and H1 versus changes in mean domain distances (x-axis). Gene distribution is shown as Kernel density plots on the right.



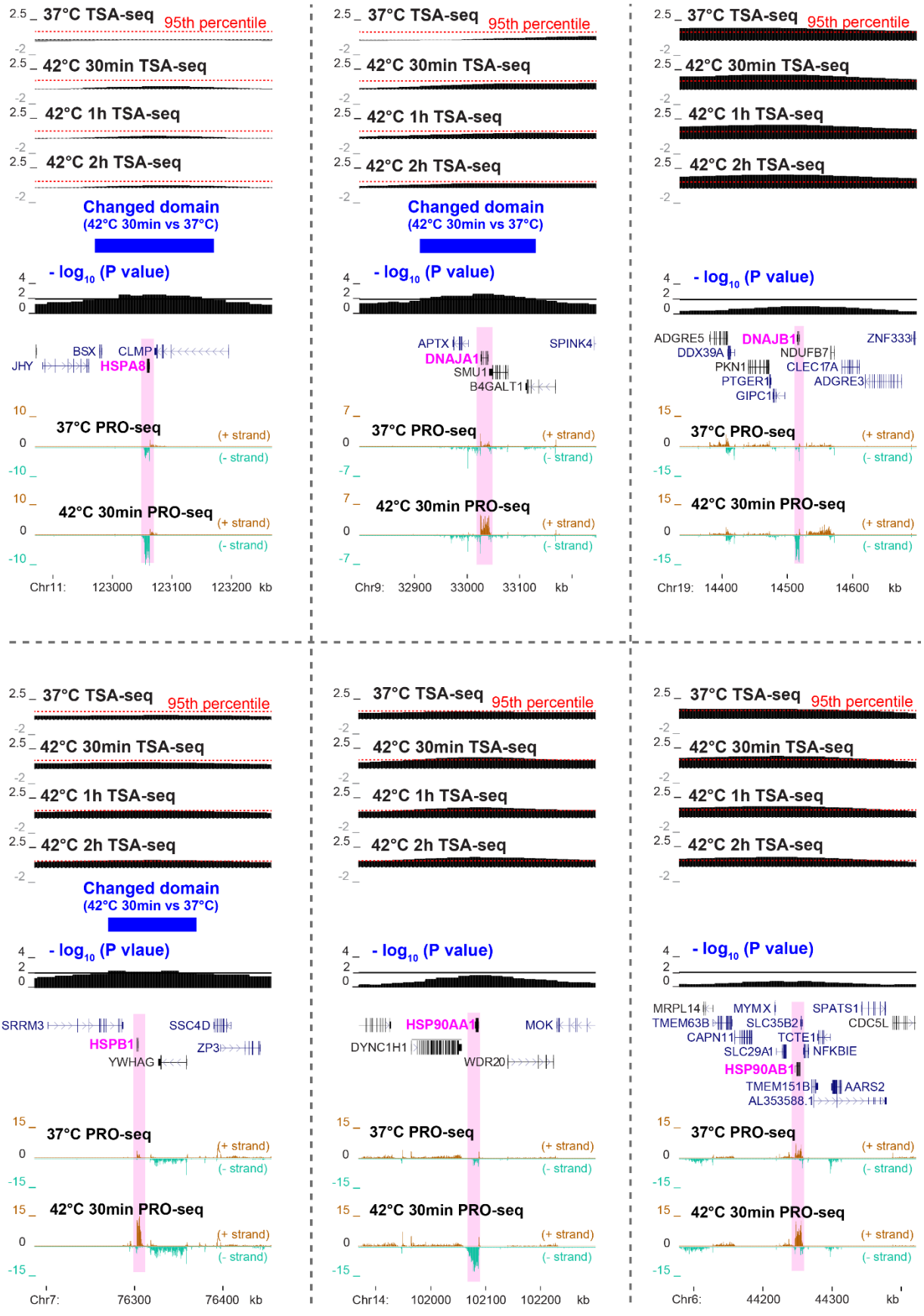
Supplemental Fig S8. Cell type differences in TSA-seq genomic percentiles of significantly changed regions. Domain mean TSA-seq enrichment score percentiles of significantly changed regions relative to SPADs (**A**) or all significantly changed regions genome-wide (**B**) in cell type pair-wise comparisons. Each dot represents one region.



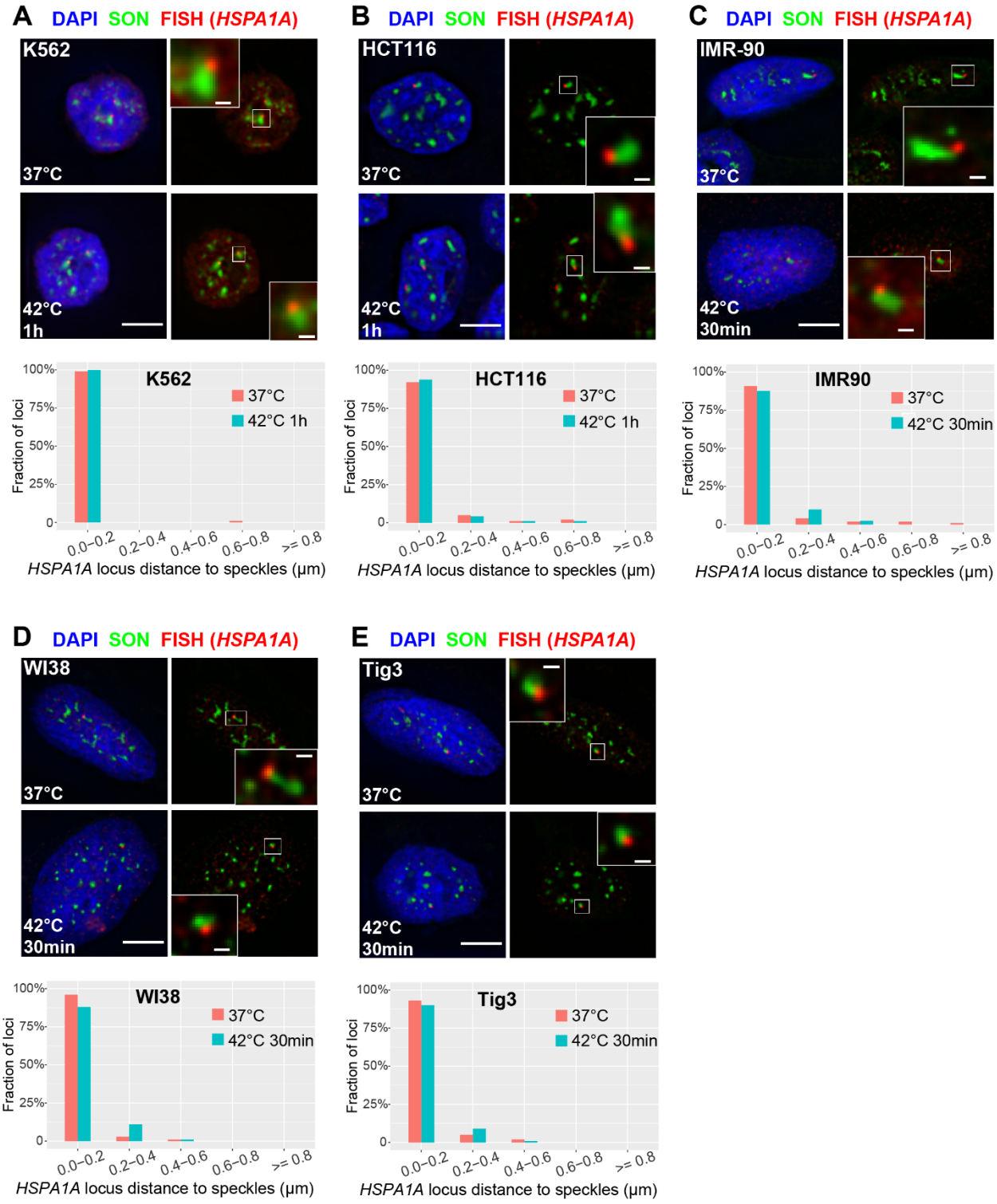
Supplemental Fig S9. Increased Collagen gene expression in HFF fibroblasts versus H1 hESCs correlates with closer relative distance to nuclear speckles in fibroblasts versus hESCs. A, C) Comparison of H1 and HFFc6 smoothed SON TSA-seq enrichment scores: top track- H1 (green), middle track- HFFc6 (brown), bottom track- p-values per bin for significance of change in scaled SON TSA-seq scores. Dashed lines (top and middle tracks) indicate "zero" values showing regions with genome-wide average number of reads. Highlighted regions (brown) below middle track containing Collagen genes (**A, B**: COL3A1, COL5A2; **C, D**: COL4A1, COL4A2) localize closer to speckles in HFFc6. **B, D)** Zoomed browser views of same highlighted regions also showing gene annotation (GENCODE v29) and RNA-seq RPM values.



Supplemental Fig S10. Genome organization relative to nuclear speckles as measured by TSA-seq is largely invariant after heat shock. A) SON immunostaining (green) in K562 cells at 37 °C or upon heat shock at 42 °C for 30 mins, 1 hr or 2hrs with DAPI (blue, right). **B)** SON TSA-seq mapping results with or without heat shock showing smoothed TSA-seq enrichment scores. **C)** 2D histograms showing correlation of smoothed SON TSA-seq enrichment scores between replicates or different conditions. Colors represent numbers of 20 kb genomic bins with given TSA-seq scores in the 2 datasets falling within given histogram 2D bin (~0.03 x 0.03 of TSA-seq score intervals). Pearson's r: the Pearson product-moment correlation coefficient.


37°C TSA-seq
95th percentile
42°C 30min TSA-seq
42°C 1h TSA-seq
42°C 2h TSA-seq
Changed domain
(42°C 30min vs 37°C)
- log10 (P value)
37°C PRO-seq
42°C 30min PRO-seq
(+ strand)
(- strand)
JHY
BSX
CLMP
HSPA8
Chr11: 123000 123100 123200 kb
APTX
SPINK4
DNAJA1
SMU1
B4GALT1
Chr9: 32900 33000 33100 kb
ADGRE5
DNAJB1
ZNF333
DDX39A
NDUFB7
PKN1
CLEC17A
PTGER1
ADGRE3
GIPC1
Chr19: 14400 14500 14600 kb
SRRM3
SSC4D
HSPB1
ZP3
YWHAG
Chr7: 76300 76400 kb
HSP90AA1
MOK
DYNC1H1
WDR20
Chr14: 102000 102100 102200 kb
MRPL14
MYMX
SPATS1
TMEM63B
SLC35B2
CDC5L
CAPN11
TCTE1
SLC29A1
NFKBIE
HSP90AB1
TMEM151B
AARS2
AL353588.1
Chr6: 44200 44300 kb

Supplemental Fig S11. Major inducible heat shock protein genes reside in unchanged SPADs or domains moving small distances to speckles. TSA-seq and PRO-seq profiles of multiple induced heat shock protein (HSP) genes in K562 cells upon heat shock. From top to bottom: smoothed SON TSA-seq enrichment score tracks of control or different heat shock time points (red dashed lines indicate 95th TSA-seq score percentiles), significantly changed domains comparing 37 °C vs 42 °C 30 mins (blue), -$\log_{10}$ (p-values of 20 kb-bins for statistical comparison of rescaled SON TSA-seq scores between 37 °C and 42 °C 30 mins), gene annotation (GENCODE) and K562 PRO-seq tracks showing normalized read count (BPM, 50 bp bins).



Supplemental Fig S12. *HSPA1A* locus deterministically positions at the periphery of nuclear speckles by microscopy in multiple cell lines. A-E) Top: 3D immuno-FISH for *HSPA1A* locus (red) plus SON

immunostaining (green) and merged channels with DAPI (blue, left, scale bar: 5 μm) in K562 (**A**), HCT116 (**B**), IMR90 (**C**), WI38 (**D**), or Tig3 (**E**) cells at 37 °C or after 42 °C heat shock. Insets: 4× enlargement of the white-boxed image area (right, scale bar: 0.5 μm). **A-E) Bottom:** Distribution comparison of distances from *HSPA1A* FISH signal to a closest nuclear speckle in K562 (**A**), HCT116 (**B**), IMR90 (**C**), WI38 (**D**), or Tig3 (**E**) cells between control and heat shock treated samples. K562: n=101 (37 °C) or 104 (42 °C 1hr). HCT116: n=101 (37 °C) or 97 (42 °C 1hr). IMR90: n=98 (37 °C or 42 °C 30 mins). WI38: n=100 (37 °C or 42 °C 30 mins). Tig3: n=100 (37 °C or 42 °C 30 mins).

Supplemental Tables

Supplemental Table S1. Data values used in this study for speckle distance calibration derived from previous immuno-FISH measurements described in our previous publication (*J Cell Biol* (2018) 217 (11): 4025–4048.)

BAC	Genome coordinates (hg38)	Mean distance to speckles (um)
RP11-634L10	chr17:81,838,939-82,011,417	0.09
RP11-479I13	chr6:31,726,514-31,941,167	0.11
RP11-264N5	chr7:100,470,712-100,665,336	0.16
RP11-1058N17	chr18:48,801,893-48,998,009	0.47
CTD-3106L12	chr2:24,775,317-24,986,884	0.5
RP11-997B19	chr17:71,701,964-71,881,248	0.81
RP11-978O5	chr2:22,703,020-22,897,142	0.97
RP11-846O11	chr18:41,032,947-41,237,108	0.98
RP11-302K17	chr10:102,058,244-102,216,677	0.25
RP11-246J15	chr1:202,111,935-202,271,377	0.36
CTD-3244P16	chr10:102,990,564-103,165,048	0.47
CTD-2503D10	chr10:103,950,393-104,169,256	0.51
RP11-729K13	chr2:30,387,390-30,582,345	0.6
RP11-53I19	chr6:23,302,021-23,446,224	0.76
RP11-543G21	chr1:199,421,727-199,594,861	0.92
RP11-1047B3	chr7:114,496,093-114,692,887	0.97

Supplemental Table S2. Parameters for exponential fitting ($y = y_0 + Ae^{R_0 x}$) of TSA-seq enrichment ratios (y) and mean speckle distances (x) using a "hybrid" method (Supplemental Methods)

K562 SON TSA	Condition A	Condition B	Condition C	Condition D	Condition E
y_0	0.27	0.35	0.24	0.38	0.28
A	9.69	6.10	7.7	5.08	6.75
R_0	-4.28	-3.53	-3.80	-3.43	-3.79

Supplemental Table S3. Cell lines used in this study

Cell line	Organism	Lineage
H1	Human	Embryonic stem cell
K562	Human	Erythroleukemia
HCT116	Human	Colorectal carcinoma (epithelial)
HFFc6	Human	Foreskin fibroblast

Supplemental Table S4. Sequences of single-molecule RNA-FISH probes used in this study

***HSPH1* smRNA-FISH probes**
agaatctgcggcatttactc
tggctgataagaaaccctgg
tccaatttgtgatggcaagc
cgtgaatccctgagacaatc
cccagtaggttttgaaaacc
gcgcaaagatatgcgcgaag
cggaagaaaaaggccggcag
cttccagttcatttcgagat
tggacaagcagcgacacaag
tggtcgctaacaacaatccg

gtgggctaatctttcatagg
ggccaacaatgttagcaacg
catctttcgacaatgatccc
gcgtcctagcacaaaagtat
ttagttcagtgacgacatcc
taattgtattaccccagtca
tcataaacgttctcagcacc
acagtcttcctaacgttact
aacttagatctgcaggggta
aaggggaactccaatactgc
atctactcattttcctaggg
gcatcactgtaagctctgaa
tacctgcttaggtatgagta
ccgcacaataacagagcaga
gtcgagggaaaggagtgact
tttgttctgatataccaccg
gatgtgactctacctttcag
ttgaccctacacacgactaa
tactaccccaaaaacctctg
aggggatagggaacaagcat

Supplemental Table S5 is a separate Excel file showing public RNA-seq datasets used for correlation with TSA-seq and for differential expression analysis between H1 and HFF cells.

Supplemental Table S6 is a separate Excel file showing results of RNA-seq data analysis and correlation TSA-seq data in the four cell lines.

Supplemental Table S7 is a separate Excel file showing all identified differentially expressed protein-coding genes between H1 and HFF cells.

Supplemental Table S8 is a separate Excel file showing results of Gene Ontology analysis for differentially expressed protein-coding genes in regions that are relatively closer to speckles versus that are not relatively closer to speckles between H1 and HFF cells.

Supplemental Table S9 is a separate Excel file showing summary of TSA-seq data used in this study.