TSA-Seq 2.0 reveals both conserved and variable chromosomal distances to nuclear speckles

Liguo Zhang[1], Yang Zhang[2], Yu Chen[1#], Omid Gholamalamdari[1], Jian Ma[2], Andrew S. Belmont[1,3,4]*

[1]Department of Cell and Developmental Biology, University of Illinois at Urbana-Champaign, Urbana, IL
[2]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA
[3]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL;
[4]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL

[#]Present address: Department of Molecular and Cell Biology, Li Ka Shing Center for Biomedical and Health Sciences, CIRM Center of Excellence, University of California, Berkeley, CA 94720, USA , Howard Hughes Medical Institute, Berkeley, CA 94720, USA

*Corresponding author: Andrew Belmont, asbel@illinois.edu

Summary:

TSA-Seq measures chromosomal distances from specific nuclear compartments genome-wide but requires ≥100 million cells. We report 10-20-fold increased sensitivity using TSA-Seq 2.0 which deliberately saturates protein-labeling but preserves distance mapping by the still unsaturated DNA-labeling. Mapping nuclear speckle distances in four cell lines reveals highly transcriptionally active, conserved speckle-associated chromosome domains but relative shifts of a small fraction of the genome that highly correlates with changes in gene expression.

31 High-resolution Hi-C has revealed a division of A and B compartments into several
32 subcompartments[1], while newer genomic methods such as SPRITE[2] and GAM[3] both point to
33 multi-way interactions occurring more frequently in the nuclear interior, specifically at nuclear
34 bodies such as nucleoli and nuclear speckles[2]. In addition to mapping contact frequencies, what
35 are needed are convenient genomic methods that directly interrogate the spatial relationship of
36 chromosomes, or even actual distances, relative to specific nuclear bodies/compartments.
37 Tyramide Signal Ampification (TSA) uses a horse-radish peroxidase (HRP)-conjugated secondary
38 antibody to generate biotin-tryamide free radicals that diffuse and then covalently react with
39 nearby macromolecules[4, 5]. TSA-Seq exploits the exponential decay gradient of tyramide
40 labeling from a point source to convert sequencing reads into a "cytological ruler", reporting on
41 mean chromosome distances from immuno-stained nuclear compartments[6]. TSA-Seq mapping
42 on speckle and lamina in K562 cells reveals a lamina-speckle axis with gradients in gene
43 expression and genomic features, suggesting a new model of nuclear organization[6]. However,
44 as originally implemented, "TSA-Seq 1.0" requires several hundred million cells.

45 TSA is typically described as labeling the tyrosines of proteins; however, we found that
46 tryamide free radicals also label DNA[6]. We used pulldown of purified DNA rather than
47 chromatin, ensuring mapping is not biased towards either genomic regions with high tyrosine-
48 rich nonhistone protein content, given that histones are a very weak substrate for the tyramide
49 labeling, or chromatin regions of higher solubility. We designed a series of TSA labeling
50 conditions with increased tyramide-biotin concentration and reaction time (Conditions A-E)
51 using anti-SON nuclear speckle TSA labeling in K562 cells (Fig. 1a, Fig. S1). Condition A
52 corresponds to TSA-Seq 1.0 conditions[6]. Nuclear speckle TSA labeling increased from Conditions
53 A to C but plateaued, with increases in background staining, from Conditions C to E (Fig. 1a&b).

54 Condition E produced higher tyramide-biotin labeling in the cytoplasm with no visible specific
55 nuclear speckle staining above the nucleoplasmic staining.

56 We hypothesized that progressive saturation of protein tyrosines causes non-specific
57 TSA labeling spreading progressively outwards from nuclear speckles, which dominated weaker
58 DNA tyramide labeling masked by this protein labeling. Moreover, under these same
59 conditions, DNA tyramide labeling remains far from saturation.

60 To test this hypothesis, we "blocked" protein tyrosine groups using high concentrations
61 of tyramide-biotin (Condition E), and then performed a second TSA using tyramide-FITC
62 (Condition A) (Fig. 1c). TSA labeling using the under-saturating Condition A produced similar
63 nuclear speckle staining after both rounds of TSA labeling (Fig. 1d, top). As predicted (Fig. 1c),
64 however, while Condition E TSA labeling produced diffuse biotin-labeling throughout the cell, a
65 2nd TSA labeling (Condition A) revealed nuclear speckle-specific labeling in K562 (Fig. 1d, bottom)
66 and HFFc6 cells (Fig. S2).

67 We estimated average tyramide-biotin labeling increased from ~1 biotin /200 kb using
68 Condition A to ~ 1 biotin/7.5 kb using Condition E (Fig. S3), predicting ~1 biotin/0.94 kb over the
69 ~8 fold-enriched speckle TSA-Seq peaks. Sonicating DNA to ~100-600 bps ensures few DNA
70 fragments with multiple biotins, maintaining linearity between TSA labeling and pulldown read
71 number.

72 Speckle TSA-Seq maps generated from Conditions A-E were both qualitatively and
73 quantitatively similar (Fig. 1e). Converting TSA-Seq signals to predicted nuclear speckle
74 distances using exponential fitting of immuno-FISH data (Supplementary Table 1 and 2,
75 Methods) demonstrated similar distances estimated from TSA-Seq data across Conditions A-E
76 (Fig. 1e, Fig. S4a). Distance residuals across the genome between Conditions A and E were

77 mostly smaller than the microscopy diffraction limit of ~0.25 µm (Fig. 1e-f, Fig. S4). Conditions
78 A, C, and E also produced similar SON TSA-Seq maps in HCT116 cells (Fig. S5).

79 Thus using Condition E we increased pulldown DNA yields by 10-20 fold relative to TSA-
80 Seq 1.0 (Condition A) (Fig. 1g, Fig. S5d), reducing required cell numbers to ~10-15 million (to
81 obtain ~5 ng pulldown DNA), without significant quantitative changes in nuclear speckle
82 distance estimations.

83 Previously, we defined chromosome regions with top 5% nuclear speckle TSA-Seq scores
84 as Speckle Associated Domains (SPADs) as they were near-deterministically localized near
85 nuclear speckles, with mean estimated speckle distances of <0.32 µm and ~100% of alleles
86 positioned within 0.5 µm of speckles by FISH[6]. To test whether SPADs are conserved among
87 different cell types, we applied SON TSA-Seq 2.0 to four cell lines (Supplementary Table 3, Fig.
88 1a, Fig. S5a, S6), each with two biological replicates. Indeed, SPADs remained largely conserved
89 among all 4 cell types (Fig. 2a-b, Fig. S7), particularly considering the ~5 percentile differences
90 between biological replicates for regions with top 10% TSA-Seq scores (Fig. S8).

91 To correlate SPADs with gene expression, we divided human protein-coding genes into
92 20 groups according to their SON TSA-Seq scores in each cell line. SPADs were dramatically
93 enriched in the top 5% expressed genes (Fig. 2c) and had the highest FPKM values in all 4 cell
94 lines (Fig. S9), suggesting SPADs were correlated with genomic regions of constitutively high
95 gene expression. Also, SPADs were enriched in housekeeping genes compared to non-
96 housekeeping genes (Fig. 2d).

97 We next identified chromosome domains 100 kb or larger with significantly different
98 relative positions to nuclear speckles in pair-wise comparisons between cell lines. We did this
99 by merging 5 or more adjacent 20-kb bins which each showed changed TSA-Seq scores with a p-

100 value threshold of 0.01. Comparing H1 and HFFc6, we identified 494 domains localized closer to
101 speckles in HFFc6 (408 kb average size) and 367 domains (396 kb average size) closer to speckles
102 in H1 (e.g. Fig. 2e, highlighted).

103 Chromosome regions closer to nuclear speckles in one cell line versus the other typically
104 showed one or more genes with higher relative gene expression levels in this cell line (e.g. Fig.
105 2f), including many examples of genes with tissue-specific expression- for example, Collagen
106 genes in fibroblasts (Fig. 2f, Fig. S10). Genome-wide analysis confirmed this marked expression
107 bias (Fig. 2g): of 628 genes with log2-fold changes >=1, 379 genes >= 2, and 248 genes >= 3
108 located within chromatin domains closer to speckles in H1 versus HFFc6, 91%, 94%, and 96%,
109 respectively, show higher expression in H1; of 307 genes with log2-fold changes >= 1, 161
110 genes >=2, and 94 genes >=3 located within domains closer to speckles in HFFc6 versus H1, 66%,
111 77%, and 83%, respectively, show higher expression in HFF. Comparing H1 and HCT116
112 revealed a similar expression bias (Fig. S11).

113 For all differentially expressed (DE) protein-coding genes in HFF versus H1 cells, we
114 found ~9-fold more genes with significantly higher expression in regions that did not change
115 their positions versus regions that did change their positions relative to nuclear speckles (Fig.
116 S12a, d). As a group, DE genes located in chromatin domains that changed their positions
117 relative to nuclear speckles showed a moderate shift in their distribution towards higher fold-
118 changes in expression, although there was extensive overlap in expression changes with DE
119 genes in chromatin domains that did not change positions (Fig. S12b, e). Gene ontology (GO)
120 analysis suggested DE genes in regions closer to speckles in one cell line were related to cell type
121 specific functions, including extracellular matrix organization and collagen related functions in

122 HFFc6 fibroblasts (Fig. S12c) and homophilic cell adhesion related to stem cell colony formation
123 and maintenance[7] in H1 hESCs (Fig. S12f).

124 In conclusion, we used protein super-saturation labeling conditions to increase the
125 sensitivity of TSA-Seq, allowing mapping of 3D genome organization relative to nuclear speckles
126 using 10-15 million cells versus 200-400 million cells. Whereas the TSA-Seq 1.0 procedure
127 required sequential staining of several batches of cells over 1-2 months, our new TSA-Seq 2.0
128 procedure required 10-20-fold fewer cells performed in a single, one-week staining. Cell culture
129 costs were reduced commensurately- for example, for hESCs from ~$2000 to ~$200 per
130 replicate.

131 These reductions in time and cost allowed us to apply TSA-Seq 2.0 to compare nuclear
132 relative distances to nuclear speckles genome-wide in 4 cell lines. This mapping revealed a
133 conservation of SPADs, which in all cell lines were associated with unusually high gene
134 expression levels. More surprisingly, pair-wise comparisons of cell lines showed remarkable
135 similarities in relative nuclear speckle distances for ~90% of the genome. Moreover, for the
136 several hundred chromatin domains that did show significant changes in TSA-Seq scores, these
137 changes typically were modest, ranging in scaled TSA-Seq scores (1 – 100) from a minimum of 11
138 to a maximum of 26 (mean 13, median 13) for the HFFc6 and H1 comparison. Yet these changes
139 highly correlated with increased (decreased) gene expression levels for regions that were closer
140 to (further from) nuclear speckles.

141 The improved sensitivity of TSA-Seq 2.0 will greatly facilitate extension of TSA-Seq to
142 mapping of additional nuclear compartments across a range of cell types and conditions.
143 Correlating changes in chromosome positioning relative to nuclear compartments/bodies using

144 TSA-Seq 2.0 with changes in gene expression, DNA replication timing, and other genomic

145 features should help reveal general principles of nuclear genome organization and function.

146

147 **Methods**

148 **Cell culture**

149 K562 cells were obtained from ATCC and cultured according to ENCODE Consortium
150 protocol (http://genome.ucsc.edu/ENCODE/protocols/cell/human/K562_protocol.pdf). H1-ESC
151 (WA01), HCT116, HFF-hTert-clone 6 cells were obtained through 4D Nucleome Consortium and
152 cultured according to 4DN SOPs (https://www.4dnucleome.org/cell-lines.html). Briefly, K562
153 cells were cultured in RPMI-1640 medium supplemented with 10% FBS (Sigma F2442) and 1%
154 100X Antibiotic-Antimycotic (GIBCO). Cells were seeded at 0.1 million/mL density and passaged
155 or harvested at around 0.75 million/mL density. H1 cells were cultured in Matrigel (Corning
156 354277, Lot #7128002)- coated flasks in mTeSR medium (STEMCELL Tech 85850). Cells were
157 passaged (or harvested) at optimal density and colony size (see SOP) every 4 or 5 days with 1 in
158 10 to 1 in 20 splits by digesting colonies into 50-200 µm aggregates using ReLeSR (STEMCELL
159 Tech 05872). HFFc6 cells were cultured in DMEM medium supplemented with 20% heat-
160 inactivated FBS (VWR 97068-091, Lot #035B15). Cells were passaged (or harvested) at 70%-80%
161 confluency without significant drop of mitotic cell ratio with 1 in 2 splits using Trypsin-EDTA
162 (0.05%) (Fisher Sci 25300054). HCT116 cells were cultured in McCoy's 5A Medium supplemented
163 with 10% heat-inactivated FBS (VWR 97068-091, Lot #035B15). Cells were passaged (or
164 harvested) at 70%-80% confluency (around 0.4 million cells/cm^2) and seeded at $4-5 \text{ x } 10^4$
165 cells/cm^2 using Trypsin-EDTA (0.05%) (Fisher Sci 25300054).

166 **Coverslip TSA staining**

167 K562 cells were plated on poly-L-lysine (Sigma P4707, 70,000-150,000 M.W., 0.01% w/v)
168 coated coverslips (Fisher Sci 12-545-81) with 0.3-0.5 mL at 0.4-0.7 million/mL cell density. Cells
169 were cultured for 30 mins for attachment. HFFc6 cells were plated on coverslips 1 or 2 days

170 before experiments and harvested at ~80% confluency. Cells were fixed with 1.6% freshly-made
171 paraformaldehyde (PFA) (Sigma P6148) in PBS at room temperature (RT) for 20 mins. Cells were
172 then permeabilized with 0.5% Triton X-100 (Sigma T8787) in PBS (0.5% PBST) at RT for 30 mins,
173 treated with 1.5% H_2O_2 in PBS at RT for 1 hr to quench endogenous peroxidases, and rinsed
174 with 0.1% Triton X-100 (Sigma T8787) in PBS (0.1% PBST) 3x at RT. Cells were blocked with 5%
175 normal goat serum (Sigma G9023) in 0.1% PBST (GS blocking buffer) at RT for 1 hr and then
176 incubated with rabbit anti-SON polyclonal antibody[6] (Pacific Immunology Corp, custom-raised)
177 1:2000 in GS blocking buffer at RT for 5 hrs. Cells were then washed with 0.1% PBST at RT 3 x 5
178 min and incubated with HRP conjugated goat anti-rabbit polyclonal antibody (Jackson Immuno
179 111-035-144) 1:1000 in GS blocking buffer at RT for 5 hrs or at 4 °C for 10-12 hrs. Cells were then
180 washed with 0.1% PBST at RT for 3 x 5 min and subject to TSA labeling.

181 For Condition A, the reaction solution is 50% sucrose (w/v), 1/10000 tyramide-biotin or
182 tyramide-FITC (v/v), and 0.0015% H_2O_2 (v/v) in PBS. Tyramide-biotin was prepared as previously
183 described[6]. Tyramide-FITC was prepared according to an online protocol[8]. Labeling is done at RT
184 for 10 mins. For Condition E, the reaction solution is 50% sucrose, 1/300 tyramide-biotin and
185 0.0015% H_2O_2 in PBS. Labeling is done at RT for 30 mins. For both conditions, 500 uL reaction
186 solution was applied per coverslip.

187 For "two-rounds" of TSA labeling, after the 1st-round of TSA, cells were washed with
188 0.1% PBST at RT for 3 x 5 mins and then subject to a 2nd-round of TSA. After TSA labeling, cells
189 were washed with 0.1% PBST at RT for 3 x 5 mins, stained with Streptavidin-Alexa Fluor 594
190 (Invitrogen) 1:200 and goat anti-rabbit – Alexa Fluor 647 (Jackson Immuno) 1:200 in GS blocking
191 buffer at RT for 2 hrs or at 4 °C for 10-12 hrs, and then washed with 0.1% PBST at RT for 3 x 5
192 mins. Coverslips were mounted in DAPI containing, anti-fading media (0.3 µg/ml DAPI (Sigma-

193 Aldrich)/10% w/v Mowiol 4-88(EMD Millipore)/1% w/v DABCO (Sigma-Aldrich)/25% glycerol/0.1
194 M Tris, pH 8.5).

195 **TSA-Seq**

196 The TSA-Seq procedure was modified from our previous publication[6].

197 ***For suspension cells (K562)***, cells were fixed by adding 8% freshly made PFA in PBS to
198 reach a final concentration of 1.6% and incubated at RT for 20 min. Aldehyde groups were
199 quenched by adding 1.25M (10x) glycine in PBS and mixing at RT for 5min. Cells were
200 permeabilized with 0.5% PBST at RT for 30 mins, centrifuged at 116 g, and re-suspended in PBS.
201 H_2O_2/PBS was added to reach a final H_2O_2 1.5% concentration to quench endogenous
202 peroxidases in a volume of 1 mL per 3 million cells; the cell suspension was incubated by slowly
203 nutating at RT for 1 hr (open tubes 2 or 3 times during the incubation to release the generated
204 gas). Cells were rinsed 3x with 0.1% PBST, blocked with 5% normal goat serum (Sigma G9023) in
205 0.1% PBST (GS blocking buffer) in a volume of 1 mL / 10 million cells at RT for 1 hr, and then
206 incubated with rabbit anti-SON polyclonal antibody[6] (Pacific Immunology Corp, custom-raised)
207 1:2000 in GS blocking buffer at 1mL / 10 million cells at 4 °C for 20-24 hrs. Cells were then
208 washed with 0.1% PBST at RT for 3 x 5 mins and incubated with HRP conjugated goat anti-rabbit
209 polyclonal antibody (Jackson Immuno 111-035-144) 1:1000 in blocking buffer in a volume of 1
210 mL / 10 million cells at 4 °C for 20-24 hrs. Cells were washed with 0.1% PBST for 3 x 5 mins,
211 washed with PBS for 5 mins at RT, and then subjected to TSA labeling. Cells were resuspended in
212 50% sucrose/PBS and then the same volume of 50% sucrose/PBS containing tyramide-biotin and
213 hydrogen peroxide was added to reach the final concentrations for a specific labeling condition
214 (Fig. S1). The final volume of reaction solution was 1 mL per 10 million. Cells were gently nutated
215 at RT during the TSA labeling time specific to each condition (Fig. S1). Cells were then washed at

216 RT with 0.1% PBST 3 x 5 mins and with PBS for 5 min. For each sample, a small portion of cells
217 were attached to a coverslip for anti-biotin and anti-SON staining to visualize the TSA labeling.
218 Remaining cells were pelleted and either immediately subjected to genomic DNA isolation or
219 stored at -80 °C for later DNA isolation. K562 cells were lysed with high T-E buffer (10 mM Tris
220 and 10 mM EDTA, pH 8.0) containing 0.5% SDS and 0.2 mg/mL Proteinase K (NEB P8107S). All
221 centrifugations prior to TSA labeling were low-speed at 116 – 130 g for 5 – 10 mins to preserve
222 cell structure.

223 ***For attached cells (H1, HFFc6, HCT116)***, cells were grown in tissue culture flasks and
224 fixed by quickly pouring away growth media, adding freshly made 1.6% PFA in PBS, and
225 incubating at RT for 20 mins. Cells were rinsed with PBS, and then washed/permeabilized with
226 0.5% PBST at RT for 3 x 5 mins. Free aldehyde groups were quenched with 20 mM glycine in PBS
227 at RT for 3 x 5 mins. Cells were then washed with PBS and incubated with 1.5% H_2O_2 in PBS at
228 RT for 1 hr. Cells were rinsed 3x with PBS, blocked with 5% normal goat serum (Sigma G9023) in
229 PBS in a volume of 1mL per 25 cm^2 flask surface area at RT for 1 hr, and then incubated with
230 rabbit anti-SON polyclonal antibody[6] (Pacific Immunology Corp, custom-raised) 1:2000 in 0.1%
231 PBST in a volume of 1 mL per 25 cm^2 at 4 °C for 10-12 hrs. Cells were then washed with 0.1%
232 PBST at RT for 3 x 5 mins and incubated with HRP conjugated goat anti-rabbit polyclonal
233 antibody (Jackson Immuno 111-035-144) 1:1000 in 0.1% PBST in a volume of 1 mL per 25 cm^2 at
234 4 °C for 10-12 hrs. Cells were washed at RT with 0.1% PBST for 3 x 5 mins and with PBS for 5
235 mins. Cells were TSA-labeled with a working solution of 50% sucrose/PBS containing 0.0015%
236 H_2O_2 and the condition-specific tyramide-biotin concentration (Fig. S1) in a volume of 0.8mL per
237 10 cm^2 surface area. Cells were incubated at RT for the condition-specific time (Fig. S1) and then
238 washed at RT with 0.1% PBST 3 x 5 mins and with PBS for 5 mins. For each staining, a small
239 portion of attached cells were scraped off and loaded onto coverslips for anti-biotin and anti-

240 SON immunostaining to visualize the TSA-labeling. Remaining cells were washed with high T-E
241 buffer (10 mM Tris and 10 mM EDTA, pH 8.0) at RT for 5 min and lysed with high T-E buffer
242 containing 1% SDS and 0.2 mg/mL Proteinase K. Cell lysates were collected and immediately
243 subjected to genomic DNA extraction. All incubation and washing steps before cell-lysing were
244 done by gently shaking the original flasks.

245 Cells on coverslips were stained with Streptavidin-Alexa Fluor 594 (Invitrogen) 1:200 and
246 goat anti-rabbit – FITC (Jackson Immuno) 1:500 in GS blocking buffer at RT for 2 hrs or at 4 °C for
247 10-12 hrs. Cells were then washed with 0.1% PBST at RT for 3 x 5 mins. Coverslips were mounted
248 in DAPI containing, anti-fading media (0.3 µg/ml DAPI (Sigma-Aldrich)/10% w/v Mowiol 4-
249 88(EMD Millipore)/1% w/v DABCO (Sigma-Aldrich)/25% glycerol/0.1 M Tris, pH 8.5).

250 Genomic DNA was extracted by phenol/chloroform as previously described[6]. For
251 samples with low DNA concentration, glycogen (Roche) was added to a final concentration of
252 0.05 mg/mL to facilitate ethanol precipitation. Isolated DNA was fragmented to 100-600 bp
253 using a Bioruptor Pico (Diagenode) machine with a mode of 30 sec ON – 30 sec OFF. Inspection
254 of DNA gels determined the number of required sonication cycles. Biotin labelled DNA
255 fragments were isolated with streptavidin beads as previously described[6].

256 Sequencing libraries were constructed using the TruSeq ChIP Sample Prep Kit (Illumina,
257 IP-202-1012) for H1, HCT116, HFFc6 samples, or the Hyper Library Construction Kit (Kapa Bio)
258 for K562 Condition A, B, C samples, or a protocol developed in the John Lis laboratory (Cornell
259 University) for K562 Condition D, E samples. The John Lis lab protocol includes steps of end
260 repair (End-It DNA End-Repair Kit: Epicentre Biotech ER0720), 3'-end A tailing (Klenow
261 Fragment: NEB M0212S), and indexed adaptor ligation (NEXTflex ChIP-Seq Barcodes-6: Bioo Sci
262 514120; T4 DNA Ligase (Rapid): Enzymatics Inc L6030-HC-F). Libraries were amplified by 8-12

263 PCR cycles. qPCR was used to measure the concentrations of each library which were then
264 pooled at equimolar concentration for each lane and sequenced for 101 cycles from one end of
265 the fragments on a HiSeq4000 using a HiSeq4000 sequencing kit version 1 (Illumina). Fastq files
266 were generated and demultiplexed with the bcl2fastq Conversion Software (Illumina).

267 **Microscopy**

268 3D optical sections were acquired with 0.2-µm z-steps using an Applied Precision
269 Personal DeltaVision microscope system and DeltaVision SoftWoRx software (GE Healthcare)
270 with a 60× oil objective (NA 1.4) and a CoolSNAP HQ2 charge-coupled device camera, or using a
271 DeltaVision OMX microscope system and DeltaVision SoftWorx software with a 100× (NA 1.4)
272 objective lens and an Evolve 512 Delta EMCCD Camera using the wide-field light path. Image
273 deconvolution and registration were done with SoftWoRx. Image intensity was measured with
274 FIJI software (ImageJ, NIH) using the Plot Profile function and normalized by exposure time and
275 the % transmitted exciting light. Normalized intensity plots were generated using OriginPro 2018
276 software (OriginLab). All figure panel images were prepared using FIJI and Illustrator CC (Adobe).

277 **Dot blot and DNA biotinylation estimation**

278 Biotinylation levels of DNA were assayed by dot blot prior to biotin-avidin pulldown of
279 DNA. 5-fold serial dilutions of a biotin end-labeled 250 bp fragment were used as biotin
280 concentration standards. The PCR product was produced from *Drophophila* DNA cloned within a
281 BAC (BACR48E12) using primers: GAAACATCGC/iBiodT/GCCCATAAT (forward) and
282 AGAAGCAGCTACGCTCCTCA (reverse) resulting in 1 biotin per fragment. These PCR standards
283 were combined with control, unbiotinylated, sonicated K562 genomic DNA (100-600 bp) to
284 reach a final DNA concentration the same as the test DNA sample (400 or 600 ng/uL) in order to
285 eliminate concentration effects on the crosslinking of DNA to membranes.

286 DNA was spotted with 1 or 1.5 uL (same volume for all standards and samples in a
287 specific experiment), twice onto a nitrocellulose membrane (0.45 µm; Bio-Rad), and were UV
288 cross-linked to the membrane (0.24J; UV Stratalinker 2400; Agilent Technologies). The
289 membrane was blocked with SuperBlock (TBS) blocking buffer (Thermo) with 0.05% Tween-20
290 (Fisher) at RT for 1 h, and then incubated with streptavidin-HRP (Invitrogen 43-4323) diluted
291 1:10000 in blocking buffer at 4 °C for 1-3 hr or overnight. The membrane was then washed with
292 0.05% Tween-20 in TBS at RT for 6 x 5 mins with rigorous shacking. The membrane was treated
293 with SuperSignal West Femto chemiluminescent substrate (Thermo) and developed with HyBlot
294 CL film (Denville) or in an iBright machine (Invitrogen). Sample biotinylation level (kilobases of
295 DNA with one biotin labeling on average) was calculated following the equation:

296 $$\frac{1}{l_{std}} \times c_{std} = \frac{1}{l_{smp}} \times c_{smp}$$

297 l_{std}: standard DNA length per biotin (0.25 kb). l_{smp}: average sample DNA length (kb) per biotin
298 (to be calculated). c_{std}: concentration of the 0.25kb standard DNA that has the same signal
299 intensity with the tested sample DNA. c_{smp}: concentration of the tested sample DNA.

300 **TSA-Seq data mapping and processing**

301 We used a similar analysis pipeline as the previous paper[6] to process TSA-Seq data.
302 Briefly, we mapped the raw sequencing read to the latest human reference genome (hg38,
303 downloaded from the UCSC Genome Browser) using Bowtie2[9] (version 2.0.2) with default
304 parameters. We excluded sequence from the Chromosome Y to avoid mapping bias for K562
305 data (female). We further applied the rmdup command from SAMtools[10] to remove the
306 potential PCR duplicates in the alignment. We then used the PCR duplicates removed alignment
307 files as the input files for TSA-Seq normalization.

308 TSA-Seq normalization followed a similar scheme as used in the previous paper[6], using
309 matched pulldown and input data with the following updates: First, we no longer used sliding
310 windows. Instead, we used separate genomic bins of 20kb. Each mapped read is exclusively
311 assigned to a window according to its largest aligned position in the reference genome. Second,
312 we modified the calculation of TSA-Seq enrichment score, as described in the next paragraph.

313 N_{TSA} and N_{input} represent the original number of mapped reads in each 20kb bin in the
314 pulldown and input samples, respectively. N_{TSA} in each bin (bins without mapped reads are
315 skipped) was normalized by input read number, according to equation (1), where $Ave(N_{input})$ is
316 the average value of N_{input} across genome-wide windows calculated by dividing the genome-
317 wide sum of N_{input} by the number of bins containing mapped reads:

318 $$N'_{TSA} = \frac{N_{TSA} \times Ave(N_{input})}{N_{input}} \qquad (1)$$

319 The TSA-Seq enrichment score is then defined as the log2 ratio between N'_{TSA} and the
320 genome-wide average of N'_{TSA} ($Ave(N'_{TSA})$, calculated using the number of bins containing non-
321 zero numbers of mapped reads, as per equation 2:

322 $$\text{TSA-Seq enrichment score} = \log_2\left(\frac{N'_{TSA}}{Ave(N'_{TSA})}\right) \qquad (2)$$

323 This normalized TSA-Seq enrichment score can be considered as the log2 ratio of relative
324 enrichment or depletion of DNA in a specific bin relative to the average pulldown value.

325 For subsequent analyses, the non-overlapping 20kb binned signals were smoothed by
326 convolution using a Hanning window of length 21 (21 × 20 kb or 420 kb).

327 K562 pulldown data from all five conditions were normalized using one input constructed
328 from fragmented K562 genomic DNA. H1, HCT116 and HFFc6 data were normalized using separate
329 input libraries made from the DNA used for pulldown for each TSA-labeling experiment.

330 **TSA distance prediction and comparison**

331 To convert TSA-Seq scores to distances from speckles, previously[6] we fit loci TSA-Seq
332 signal y (here y is the TSA-Seq fold-enrichment value PRIOR to the log2 operation in equation 2)
333 and mean distances (x) from speckles measured by 3D immuno-FISH to a calibration equation
334 $y = y_0 + Ae^{R_0 x}$.

335 Here we applied a new "hybrid" method to estimate the parameters for this calibration
336 equation.

337 First, we obtained the exponential decay parameter R_0 by fitting 16 FISH measurements
338 from our previously published data[6] to the new TSA-Seq data (Supplementary Table 1). For each
339 FISH probe, we used the previously published mean cytological distance to speckles based on
340 measurements of 100 alleles[6] (16 probes, Supplementary Table 1). We used the smoothed TSA-
341 Seq data and calculated the mean TSA-Seq enrichment values over the genomic regions cloned
342 within the BACs used to generate FISH probes. We fit the 16 TSA-Seq fold-enrichment values, y,
343 and their corresponding mean speckle distances, x, to the exponential function $y = y_0 + Ae^{R_0 x}$
344 using OriginPro software (OriginLab) to obtain the exponential parameter, R_0.

345 Next, we obtained y_0 and A based on the minimum and maximum TSA-Seq fold-
346 enrichment values, y_{min} and y_{max}: $y_{min} = y_0$ ($x = x_{max}$) and $y_{max} = y_0 + A$ ($x = 0$). With the
347 parameters R_0, y_0 and A (Supplementary Table 2), we converted TSA-Seq enrichment values

348 from all 20 kb bins of smoothed TSA-Seq data into speckle distances by computing the inverse
349 equation for x in the exponential function (equation (3)):

350 $$x = \frac{1}{R_0} \ln \frac{y - y_0}{A} \qquad (3)$$

351 To compare distances derived from TSA-Seq data produced using two different TSA
352 staining conditions, we calculated their distance residuals over each of the 20 kb bins and then
353 generated histograms showing the distribution of these distance residuals, using a 0.005 µm
354 histogram binning interval.

355 **SPAD calling and cell type comparisons**

356 We ranked smoothed TSA-Seq enrichment scores over all 20 kb bins genome-wide from
357 largest to smallest excluding unmapped regions from hg38 genome. We then divided them into
358 100 equal sized groups and named them as percentiles from percentile 1 to 100 (smallest to
359 largest). All adjacent 20kb bins above the 95th percentile were merged to segment these regions
360 as SPADs.

361 To compare cell lines, we first generated a scatterplot showing the variability in
362 percentile values for all 20 kb bins comparing two biological replicates. To compare SPADs called
363 in one cell line with percentile TSA-Seq scores in other cell lines, we calculated the mean TSA-
364 Seq percentile score over each SPAD region in each of the other cell lines and plotted these
365 values in a boxplot, with dots showing all values. We used Intervene[11] (version 0.6.4) to
366 generate a 4-way Venn diagram using 20-kb binned .bed files of the four cell lines.

367 **Identification of genomic domains that show different nuclear positions relative to speckles in**
368 **different cell lines**

369 To identify genomic regions that change nuclear position relative to speckles in different
370 cell lines, we adapted a previously published method[12] that compares the variability of scores in
371 two cell lines with the variability observed in biological replicates. Essentially, the hypothesis
372 tested is whether a region is statistically different from the values of the biological replicates of
373 the two different cell lines.

374 For each dataset, we first rescaled TSA-Seq enrichment scores (20kb bin) linearly
375 between their min and max values to a new 1-100 scale based on equation (4) and rounded up
376 to integers with min assigned as 1 instead of 0.

377 $$\text{Scaled enrichment score (bin i)} = \frac{\text{TSA-Seq enrichment score (bin i)} - \text{min}}{\text{max} - \text{min}} \times 100 \qquad (4)$$

378 To reduce the influence of outliers on this rescaling, we used a large and a small
379 percentile of all ranked values as the max and min values, respectively (e.g. 99.95th and 0.05th
380 percentile for HFFc6 and H1 comparison).

381 For a pair-wise comparison between two cell lines, we rescaled the SON TSA-Seq scores
382 for two biological replicates for each cell line:

383 $$\text{TSA}_{\text{H1}_{rep1}}, \text{TSA}_{\text{H1}_{rep2}}, \text{TSA}_{\text{HFFc6}_{rep1}}, \text{TSA}_{\text{HFFc6}_{rep2}}$$

384 For example, $\text{TSA}_{\text{H1}_{rep1}}$ denotes the H1 TSA-Seq biological replicate 1 and is a row vector of N
385 values from all 20kb non-overlapping bins with mapped reads in the genome. Scatterplots show
386 near uniform data noise across the genome for both cell lines (Figure 1, in this Methods section).

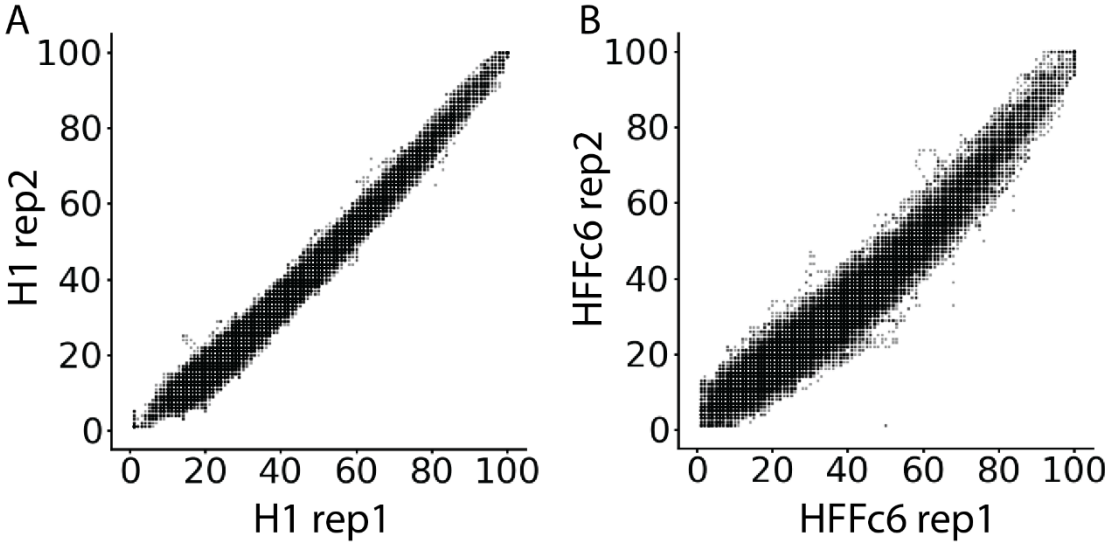


Figure 1. Scatter plots of normalized SON TSA-Seq scores between biological replicates in H1 (A) and HFFc6 (B).

We averaged the replicates for the same cell line and used the residual, Δ, between two cell lines for comparison:

$$\Delta = \frac{1}{2}\{ (\mathrm{TSA}_{\mathrm{HFFc6}_{rep1}} + \mathrm{TSA}_{\mathrm{HFFc6}_{rep2}}) - (\mathrm{TSA}_{\mathrm{H1}_{rep1}} + \mathrm{TSA}_{\mathrm{H1}_{rep2}}) \} \quad (5)$$

We define data variance as the difference between biological replicates for the same cell line. We averaged the variance between the two cell lines to be compared to construct a vector ϕ including all possible orderings with a length of 4N, where N= number of genomic bins excluding unmapped regions from hg38 genome:

$$\begin{aligned}\Phi = \frac{1}{2}\{ & (\mathrm{TSA}_{\mathrm{H1}_{rep1}} - \mathrm{TSA}_{\mathrm{H1}_{rep2}}) + (\mathrm{TSA}_{\mathrm{HFFc6}_{rep1}} - \mathrm{TSA}_{\mathrm{HFFc6}_{rep2}}), \\ & (\mathrm{TSA}_{\mathrm{H1}_{rep1}} - \mathrm{TSA}_{\mathrm{H1}_{rep2}}) + (\mathrm{TSA}_{\mathrm{HFFc6}_{rep2}} - \mathrm{TSA}_{\mathrm{HFFc6}_{rep1}}), \\ & (\mathrm{TSA}_{\mathrm{H1}_{rep2}} - \mathrm{TSA}_{\mathrm{H1}_{rep1}}) + (\mathrm{TSA}_{\mathrm{HFFc6}_{rep1}} - \mathrm{TSA}_{\mathrm{HFFc6}_{rep2}}), \\ & (\mathrm{TSA}_{\mathrm{H1}_{rep2}} - \mathrm{TSA}_{\mathrm{H1}_{rep1}}) + (\mathrm{TSA}_{\mathrm{HFFc6}_{rep2}} - \mathrm{TSA}_{\mathrm{HFFc6}_{rep1}}) \} \quad (6)\end{aligned}$$

The vector Φ serves as the null distribution against which to test single bin value difference from Δ for statistical significance. Φ can be viewed as a set of observations with random variable ϕ, which is approximately Gaussian distributed with parameters μ and σ^2 (Figure 2, in this Methods section).

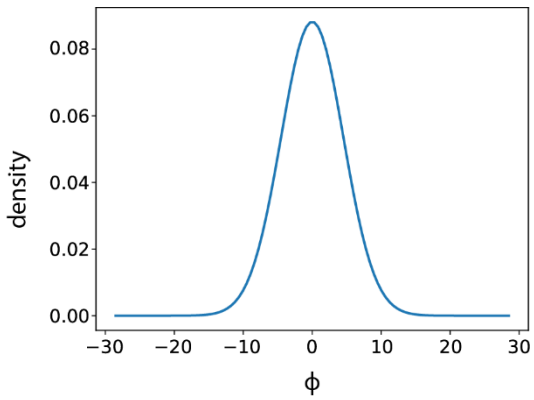


Figure 2. Density plot of ϕ

We test if a single element in Δ (a 20kb bin score) shows a significant change between the two cell lines to be compared by comparing with ϕ. We calculated p-values for both tails of the distribution separately: for equation (5), we separate the bins with values above or below 0 to obtain bins with bigger TSA scores in either HFFc6 or H1. We calculated p-values for all 20kb bins and displayed -log10 (p-value) in the genomic tracks.

To identify genomic regions that showed statistically significant changes in position, we set a cutoff p-value of 0.01 and identified all 20kb bins with p-values < 0.01 for either ordering (bigger TSA-Seq values in HFFc6 or H1). For each ordering, we merged adjacent bins whose p-values were less than 0.01 to call domains that changed location. We set a second cutoff to call only domains corresponding to 100 kb or larger.

RNA-Seq data processing and correlation with TSA-Seq data

417 We applied the ENCODE RNA-Seq processing pipeline[13] to process RNA-Seq data for
418 these cell lines (Supplementary Table 4). Briefly, we used STAR[14] (version 2.5.1b) to map the raw
419 sequencing reads to the human reference genome. The index files of the reference genome
420 used by STAR is the same that used by the ENCODE project. We downloaded the index files from
421 the ENCODE project website (URL: https://www.encodeproject.org/files/ENCFF742NER/). Next,
422 we used RSEM[15] (version 1.2.26) to quantify gene expression using the ENCODE gene annotation
423 file (URL: https://www.encodeproject.org/files/ENCFF940AZB/). The value of mRNA fragments
424 per kilobase of transcript per million mapped reads (FPKM) value for each gene reported by
425 RSEM was used for downstream analysis. Reads per million (RPM) signal tracks for each RNA-
426 Seq data set were generated using STAR with input flag "--outWigNorm RPM".

427 To correlate TSA-Seq score with gene expression, we ranked TSA-Seq enrichment scores
428 of genome-wide 20kb bins (bins without mapped reads were removed) from largest to smallest,
429 divided them into 20 equal sized groups using the cut function in R, and named them as
430 vigintiles from vigintile 1 to vigintile 20 (smallest to largest). For each protein-coding gene
431 (based on the GENCODE annotation version 24), we calculated the average TSA-Seq enrichment
432 score across the whole gene region and assigned the gene to the corresponding TSA-Seq
433 vigintile group according to the TSA-Seq enrichment score ranges for each vigintile. RNA-Seq
434 analysis and TSA-Seq correlation results were summarized in Supplementary Table 5.

435 A housekeeping gene list was download from
436 https://www.tau.ac.il/~elieis/HKG/HK_genes.txt [16] and processed as previously described[6] but
437 using hg38 RefSeq gene annotation. 3791 protein-coding genes were identified as housekeeping
438 genes, and the remaining 16541 protein-coding genes were determined as non-housekeeping
439 genes.

440 To correlate repositioned regions with expression differences between two cell lines,
441 first we identified all genes located within these genomic regions by overlapping gene and
442 region coordinates (the whole gene must be located within the region to be called). Then we
443 calculated the log2 ratio between the gene FPKM values in the two cell lines and plotted these
444 log2 ratios against the region mean scaled TSA-Seq score (max-min normalized, 1-100) change
445 between the two cell lines.

446 For differential expression analysis between H1 and HFF (RNA-Seq datasets summarized
447 in Supplementary Table 4), we used STAR[14] (version 2.5.3a) to map the raw sequencing reads
448 using the index files from the ENCODE project
449 (https://www.encodeproject.org/files/ENCFF742NER/). Next, we used htseq-count function of
450 HTSeq[17] (version 0.9.1) to count raw read number for each gene with the gencode.v24
451 annotation (https://www.encodeproject.org/files/gencode.v24.primary_assembly.annotation/).
452 To identify differentially expressed (DE) protein-coding genes between H1 and HFF, we used
453 DEseq2[18] (version 1.24.0) with thresholds for adjusted P-value of <0.01 and for fold-expression
454 change of >2-fold (Results summarized in Supplementary Table 6). For DE genes with
455 significantly higher expression in one cell line versus the other, we identified which of these
456 genes located entirely within "repositioned" domains that are located closer to nuclear speckles
457 in this cell line versus the other by overlapping gene genomic position with the coordinates of
458 the domains that showed significantly higher scaled TSA-Seq scores in this cell line. The
459 remaining DE genes with significantly higher expression in this cell line were determined as
460 differentially expressed but not located within domains that reposition closer to speckles ("non-
461 repositioned").

462 Next we calculated the mean scaled TSA-Seq scores (max-min normalized, 1-100) across
463 each of the genes and generated scatter plots to show the correlation of these TSA-Seq scores

464 between the two cell lines. We calculated log2-fold expression changes for repositioned versus
465 non-repositioned DE genes, and plotted boxplots of these expression changes for all genes in
466 each of these two categories. We used DAVID[19, 20] (version 6.8) to conduct gene ontology (GO)
467 analysis using the "GOTERM_BP_DIRECT" category (Results summarized in Supplementary Table
468 7). We compared GO terms for the repositioned and non-repositioned DE genes by plotting bar
469 plots of -log10 P-values of top 5 terms sorted by P-values.

470

471

Acknowledgements

We thank the UIUC Biotechnology center for guidance with preparation of sequencing libraries and quality control. We thank Drs. K.V. Prasanth, William Brieher, Lisa Stubbs and Huimin Zhao (UIUC, Urbana, IL) for helpful suggestions. We thank Belmont lab members for sharing reagents and providing suggestions. We thank members of the 4D-Nucleome Consortium and Belmont NOFIC center for helpful suggestions and feedback. This work was supported by National Institutes of Health grant R01 GM58460 (ASB) and U54 DK107965 (ASB, JM).

Author contributions

LZ designed experiments with ASB's guidance. LZ performed all experiments. LZ and YZ analyzed genomic data with guidance from ASB and JM. YC developed and standardized TSA-Seq 1.0 protocols and provided suggestions for development of TSA-Seq 2.0. OG contributed ideas to the "hybrid" distance-calibration mapping approach. LZ and ASB wrote the manuscript with critical suggestions from other co-authors. ASB supervised the overall study.

Data availability

All TSA-Seq data are available at the 4DN Data Portal: https://data.4dnucleome.org/browse/?experimentset_type=replicate&type=ExperimentSetReplicate&award.project=4DN&experiments_in_set.experiment_type.display_title=TSA-seq&experiments_in_set.biosample.biosource.individual.organism.name=human

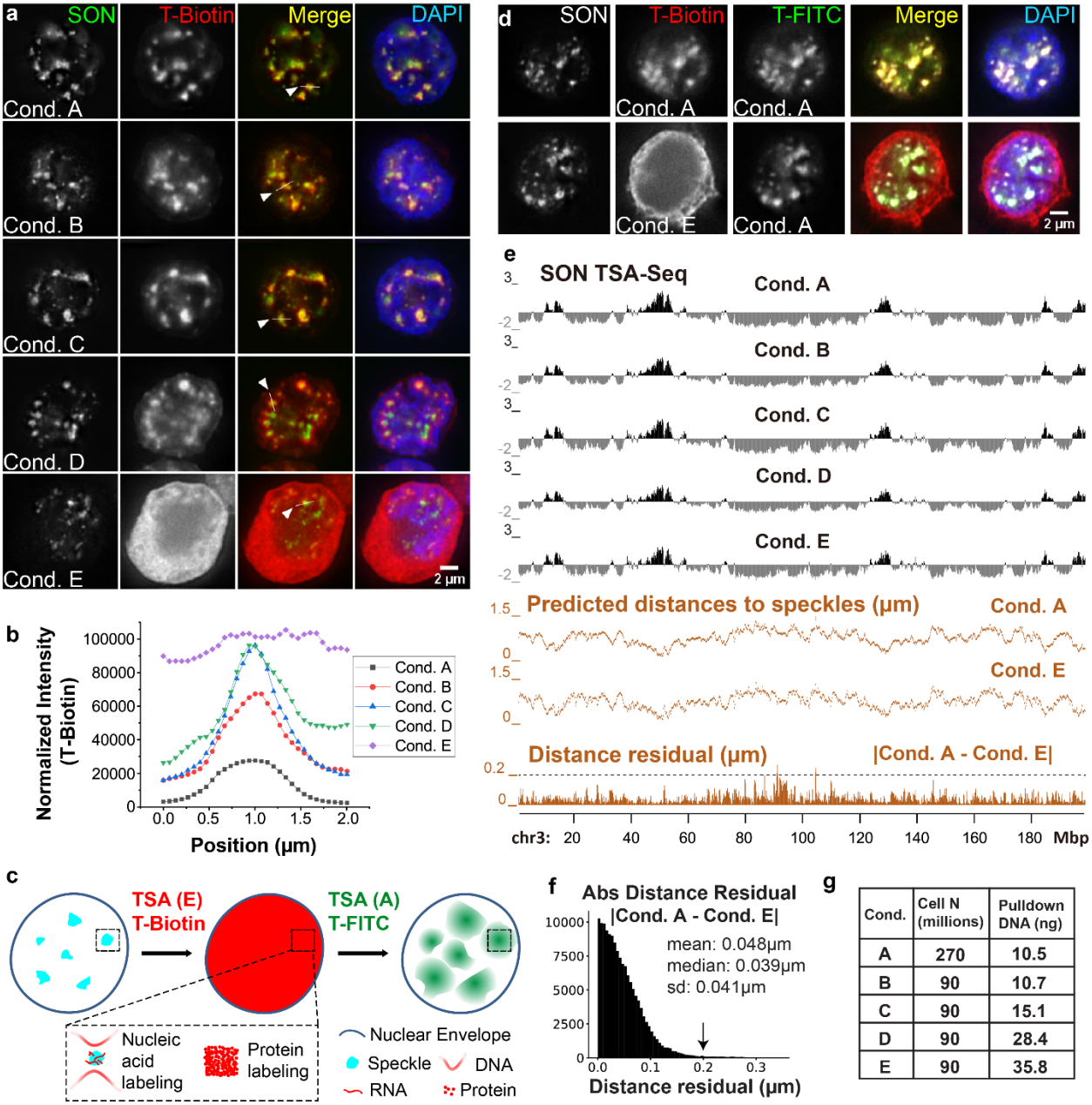
Code availability

TSA-Seq normalization software is available at https://github.com/zocean/Norma. Codes for all genomic data analyses are available at https://github.com/lgchang27/TSA-Seq-2.0-Analysis.

493 **References**

494 1. Rao, S.S. et al. *Cell* **159**, 1665-1680 (2014).
495 2. Quinodoz, S.A. et al. *Cell* **174**, 744-757 e724 (2018).
496 3. Beagrie, R.A. et al. *Nature* **543**, 519-524 (2017).
497 4. Bobrow, M.N. et al. *Journal of Immunological Methods* **125**, 279-285 (1989).
498 5. Raap, A.K. et al. *Human Molecular Genetics* **4**, 529-534 (1995).
499 6. Chen, Y. et al. *The Journal of cell biology* (2018).
500 7. Pieters, T. & van Roy, F. *J Cell Sci* **127**, 2603-2613 (2014).
501 8. http://wiki.xenbase.org/xenwiki/index.php/Flourescin_Tyramide_Synthesis.
502 9. Langmead, B. & Salzberg, S.L. *Nature methods* **9**, 357-359 (2012).
503 10. Li, H. et al. *Bioinformatics* **25**, 2078-2079 (2009).
504 11. Khan, A. & Mathelier, A. *BMC Bioinformatics* **18**, 287 (2017).
505 12. Peric-Hupkes, D. et al. *Molecular cell* **38**, 603-613 (2010).
506 13. https://github.com/ENCODE-DCC/rna-seq-pipeline.
507 14. Dobin, A. et al. *Bioinformatics* **29**, 15-21 (2013).
508 15. Li, B. & Dewey, C.N. *BMC Bioinformatics* **12**, 323 (2011).
509 16. Eisenberg, E. & Levanon, E.Y. *Trends in genetics : TIG* **29**, 569-574 (2013).
510 17. Anders, S. et al. *Bioinformatics* **31**, 166-169 (2014).
511 18. Love, M.I. et al. *Genome biology* **15**, 550 (2014).
512 19. Huang da, W. et al. *Nature protocols* **4**, 44-57 (2009).
513 20. Huang da, W. et al. *Nucleic acids research* **37**, 1-13 (2009).

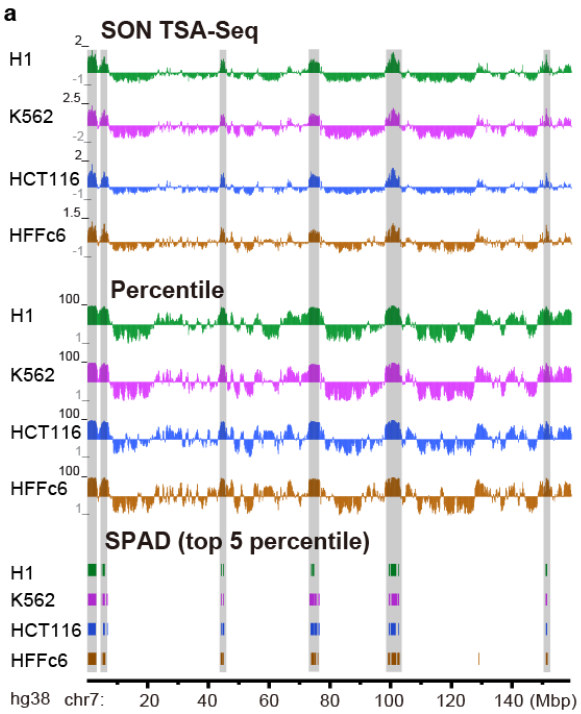
514

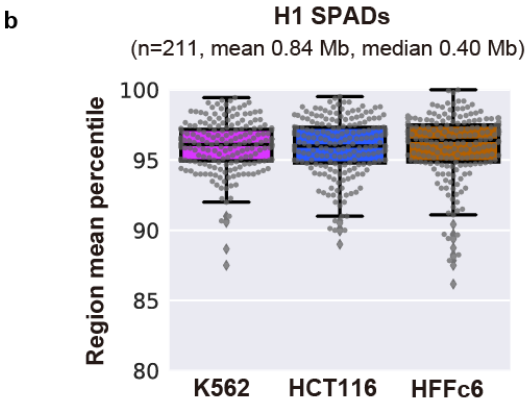


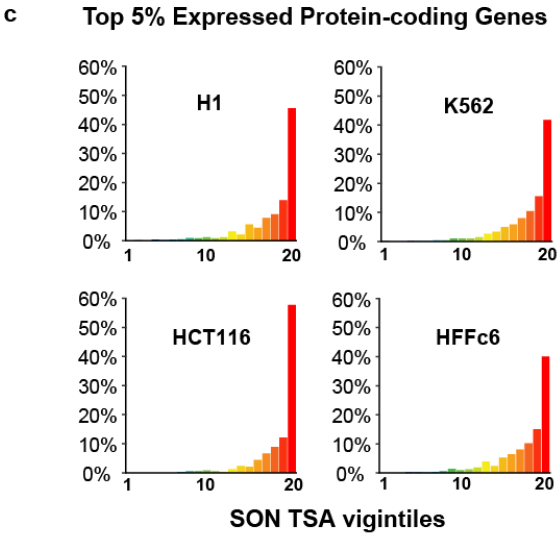
Cond.	Cell N (millions)	Pulldown DNA (ng)
A	270	10.5
B	90	10.7
C	90	15.1
D	90	28.4
E	90	35.8

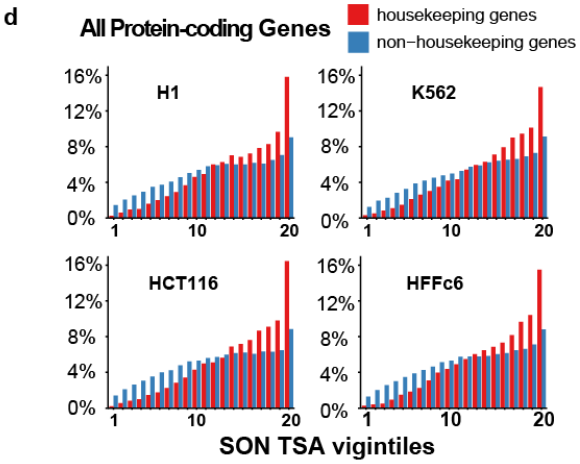
Figure 1. TSA-Seq 2.0 development and validation. a) TSA Conditions A-E show varying nuclear speckle specificity: SON immunostaining of speckles (green), strepavidin tyramide-biotin staining (red), merged channels, and DNA (DAPI, blue) in K562 cells; **b)** Tyramide-biotin intensities along line profiles spanning nuclear speckles in **a)** for Conditions A-E; **c)** Schema

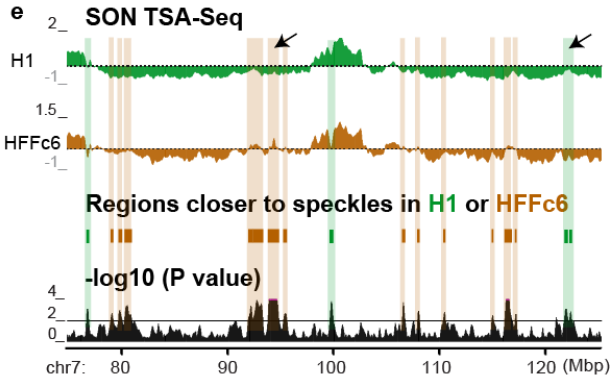
predicting results from following Condition E tyramide-biotin TSA staining with Condition A tyramide-FITC TSA staining, assuming Condition E saturates protein but not DNA tyramide-labeling; **d)** Experimental results for schema in **c)**. Top row: control showing two consecutive rounds of Condition A (non-saturating) TSA-labeling using tyramide-biotin and then tyramide-FITC. Bottom row: Same as top row but using Condition E for 1st TSA-labeling. SON immunostaining (grey), tyramide-biotin (red), tyramide-FITC (green), merged channels, plus DAPI (blue); **e)** SON TSA-Seq mapping results for Conditions A-E showing TSA-Seq enrichment scores (black tracks), estimated speckle distances (Conditions A and E, middle orange tracks), and residuals (absolute magnitude) between Conditions A and E distances (bottom, orange track); **f)** Histogram of distance residuals in **e)**: number (y-axis), residual value (x-axis); g**)** Cell numbers used and pulldown DNA yields for TSA Conditions A-E.

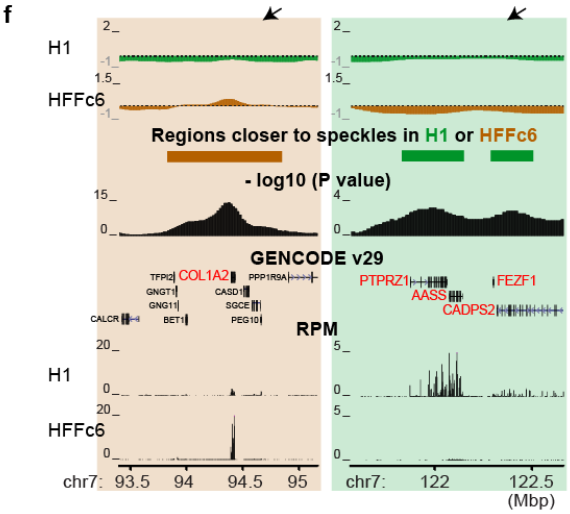

a
SON TSA-Seq
H1
K562
HCT116
HFFc6
Percentile
SPAD (top 5 percentile)
hg38 chr7: 20 40 60 80 100 120 140 (Mbp)


b
H1 SPADs
(n=211, mean 0.84 Mb, median 0.40 Mb)
Region mean percentile
K562
HCT116
HFFc6


c
Top 5% Expressed Protein-coding Genes
H1
K562
HCT116
HFFc6
SON TSA vigintiles


d
All Protein-coding Genes
housekeeping genes
non-housekeeping genes
H1
K562
HCT116
HFFc6
SON TSA vigintiles


e
SON TSA-Seq
H1
HFFc6
Regions closer to speckles in H1 or HFFc6
-log10 (P value)
chr7: 80 90 100 110 120 (Mbp)


f
H1
HFFc6
Regions closer to speckles in H1 or HFFc6
- log10 (P value)
GENCODE v29
COL1A2
PTPRZ1
FEZF1
AASS
CADPS2
RPM
chr7: 93.5 94 94.5 95
chr7: 122 122.5
(Mbp)

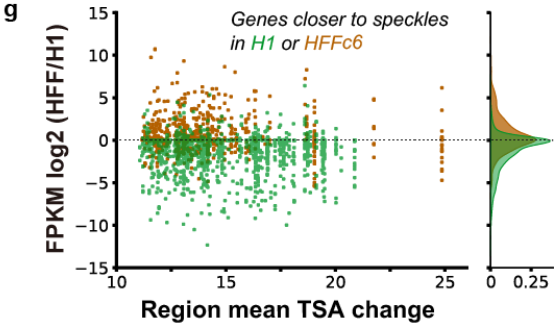
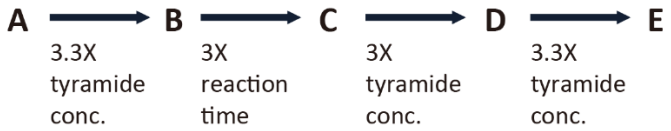

g
Genes closer to speckles in H1 or HFFc6
FPKM log2 (HFF/H1)
Region mean TSA change

Figure 2. TSA-Seq 2.0 speckle mapping in four cell lines reveals largely conserved but ~10% variable domains. a) SON TSA-Seq enrichment score (top) and SON TSA-Seq signal as a genome-wide percentile score (middle) tracks (20 kb bins), and segmented SPeckle Associated Domains (SPADs) (bottom) in 4 human cell lines (chr7). SPADs were defined as contiguous bins, each with >95th percentile score (Methods); **b)** TSA-Seq percentile score distributions of H1 SPADs in other 3 cell lines. Box plots (with data points as dots) 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), 25th percentile to lowest value within 1.5-fold of box height (bottom whisker), and outliers (diamonds); **c)** Distributions of percentages (y-axis) of top 5% expressed protein-coding genes versus SON TSA-Seq vigintiles (division of percentile scores, 0-100, into 20 bins of 5% size each) (x-axis) in 4 cell lines; **d)** 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; **e)** Comparison of H1 and HFFc6 smoothed SON TSA-Seq enrichment score tracks (top, dashed lines indicate "zero" values showing regions with genome-wide average number of reads), domains repositioned relative to speckles (highlighted, middle) and - log10 (p-values per bin for significance of change in rescaled SON TSA-Seq scores) (bottom); **f-g)** Strong bias towards increased gene expression in domains shifting closer to nuclear speckles; **f)** Zoomed view of two regions (arrows in **e**) plus gene annotation (GENCODE v29) and RNA-Seq RPM values. **g)** Changes in expression of protein-coding genes in repositioned domains: scatterplots show log2 fold-changes in FPKM ratios (y-axis) between HFFc6 and H1 versus absolute values of changes in mean scaled TSA-Seq scores

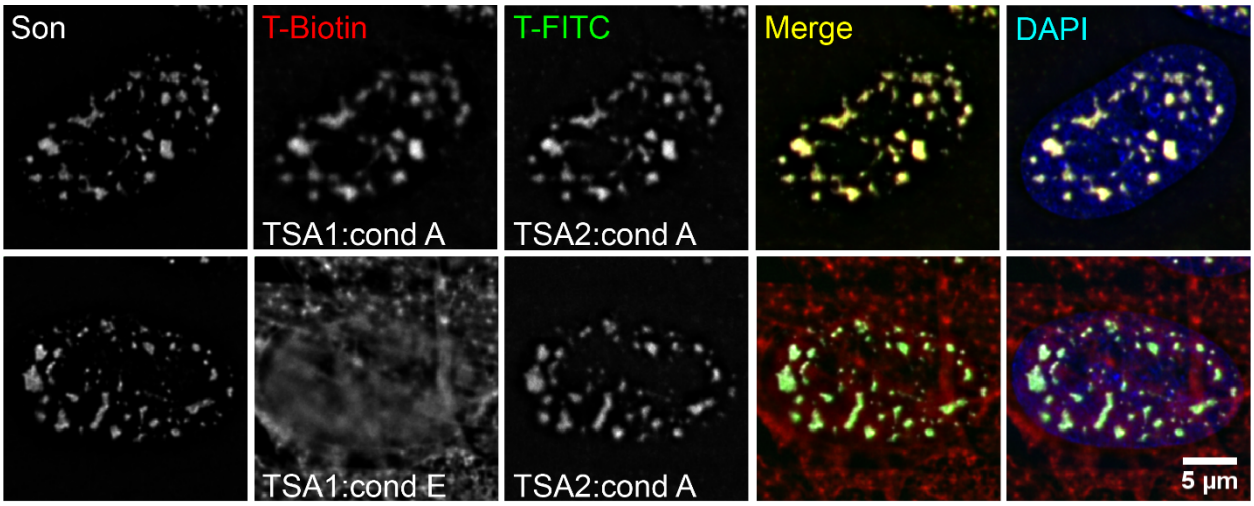
(x-axis, max-min normalized: 1-100). Green (brown) dots are genes closer to speckles in H1 (HFFc6). Gene distribution was shown as Kernel density plots on the right.

Condition:

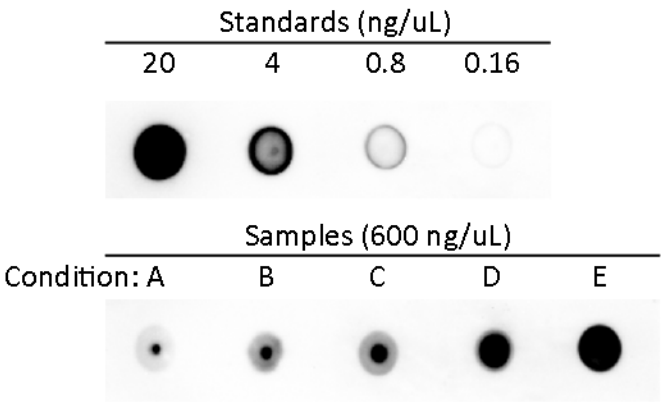


Condition	Sucrose	Tyramide-Biotin	Hydrogen peroxide	Reaction time
A	50%	1:10000	0.0015%	10 min
B	50%	1:3000	0.0015%	10 min
C	50%	1:3000	0.0015%	30 min
D	50%	1:1000	0.0015%	30 min
E	50%	1:300	0.0015%	30 min

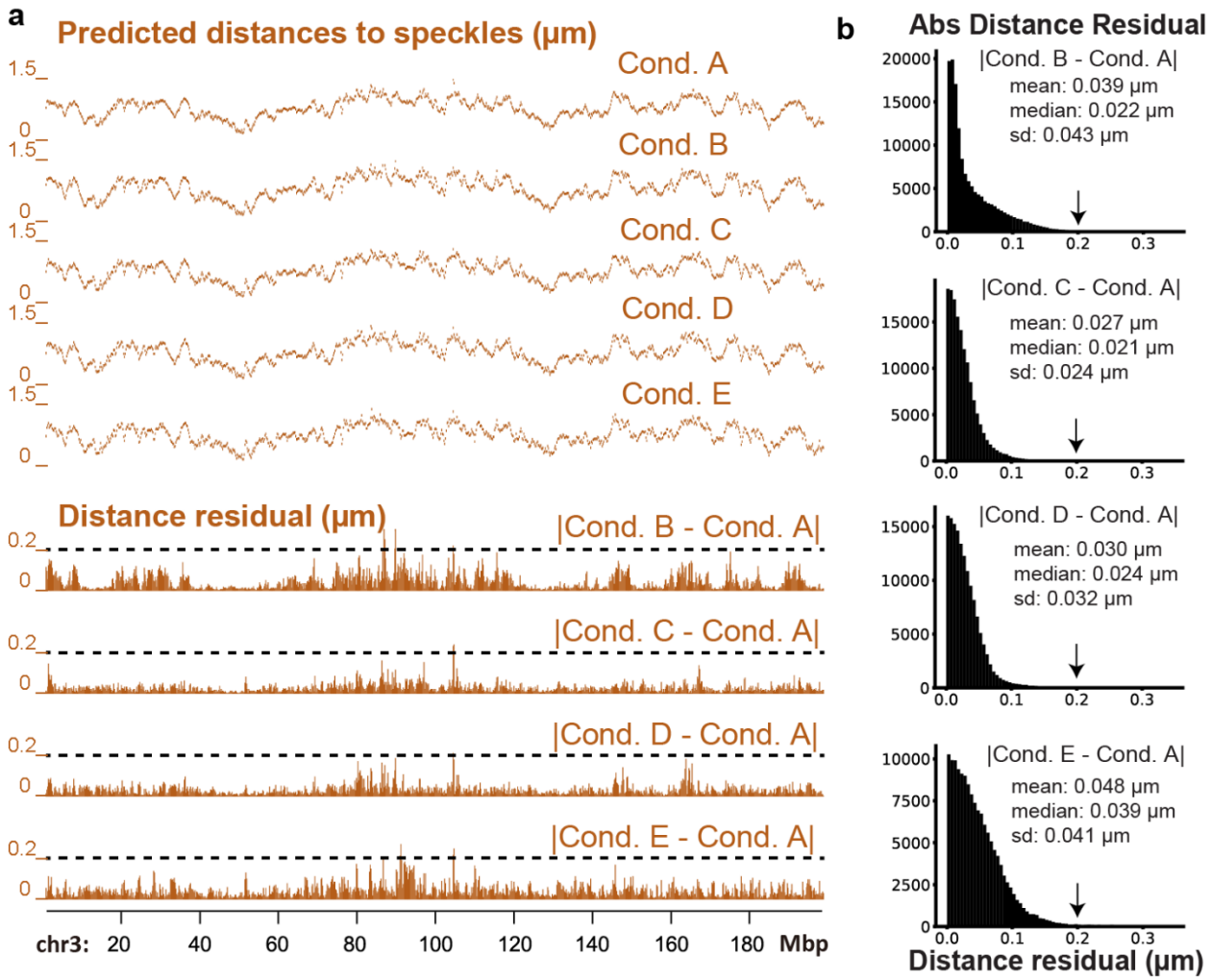
Supplementary Figure 1. Experimental design of enhanced TSA labeling conditions. All labeling conditions include 50% sucrose (w/v) and 0.0015% hydrogen peroxide (v/v). From Conditions A to E, serial increases in tyramide-biotin concentration and/or longer reaction times are applied.



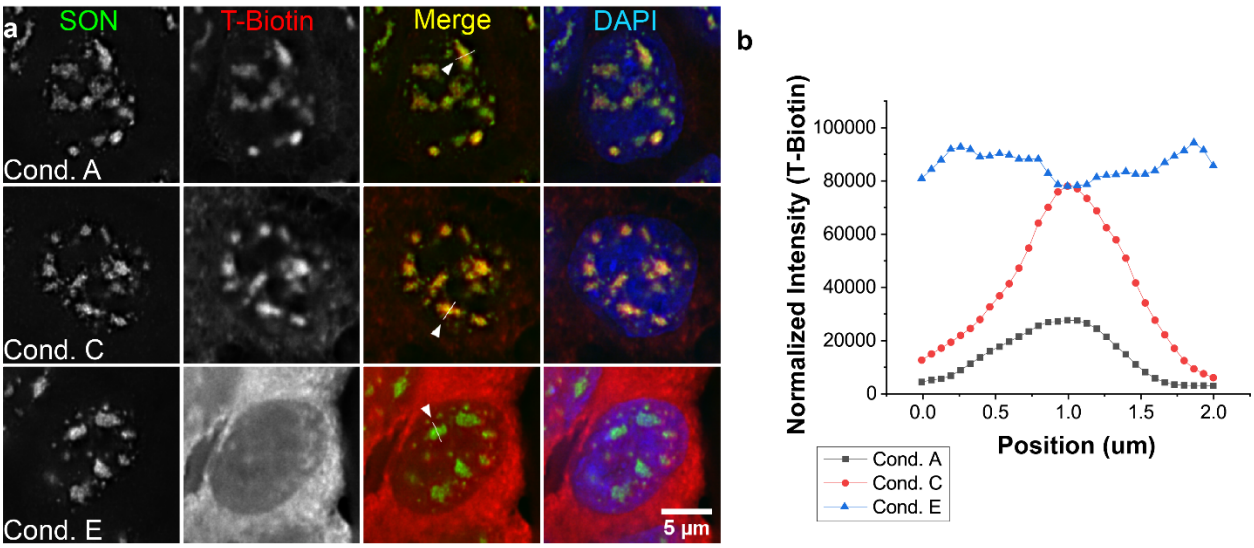
Supplementary Figure 2. Two sequential rounds of SON TSA-labeling demonstrate speckle-specific DNA TSA-labeling after previous "super-saturation" TSA staining. Top row: Two sequential Condition A (non-saturating) rounds of SON TSA-labeling reveals speckle-specific TSA-labeling for both round 1 (tyramide-biotin) 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).

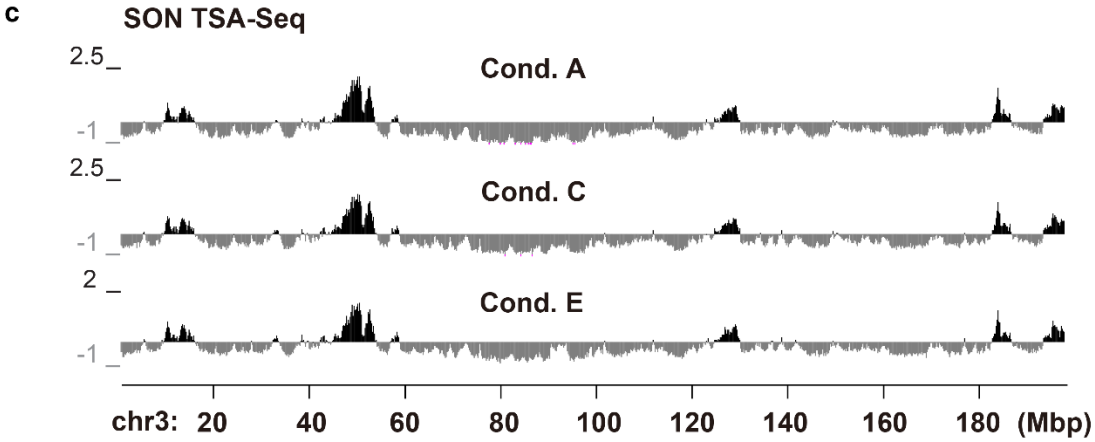


Supplementary Figure 3. Dot blot estimation of DNA-biotinylation levels after SON TSA labeling using Conditions A-E. Top: Biotinylated-DNA standards: dilutions of biotinylated PCR 250-bp product (0.16 – 20 ng/ul) containing 1 biotin / DNA molecule combined with unlabeled K562 genomic DNA (100-600 bp fragments) to a total DNA concentration of 600 ng/ul. Bottom: Sonicated genomic DNA (100-600 bp) from K562 cells after SON TSA-labeling using Conditions A-E (left to right).



Supplementary Figure 4. SON TSA Conditions A-E produce quantitatively similar genome-wide estimates of mean distances to speckles. a) Chromosome 3 tracks showing: (Top) Mean distances (um) to nuclear speckles in K562 cells estimated from SON TSA-Seq produced using Conditions A-E. (Bottom) Absolute values of distance residuals (um) 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 0.005um). Arrows mark 0.2 µm value shown as dotted horizontal lines in **a)**.



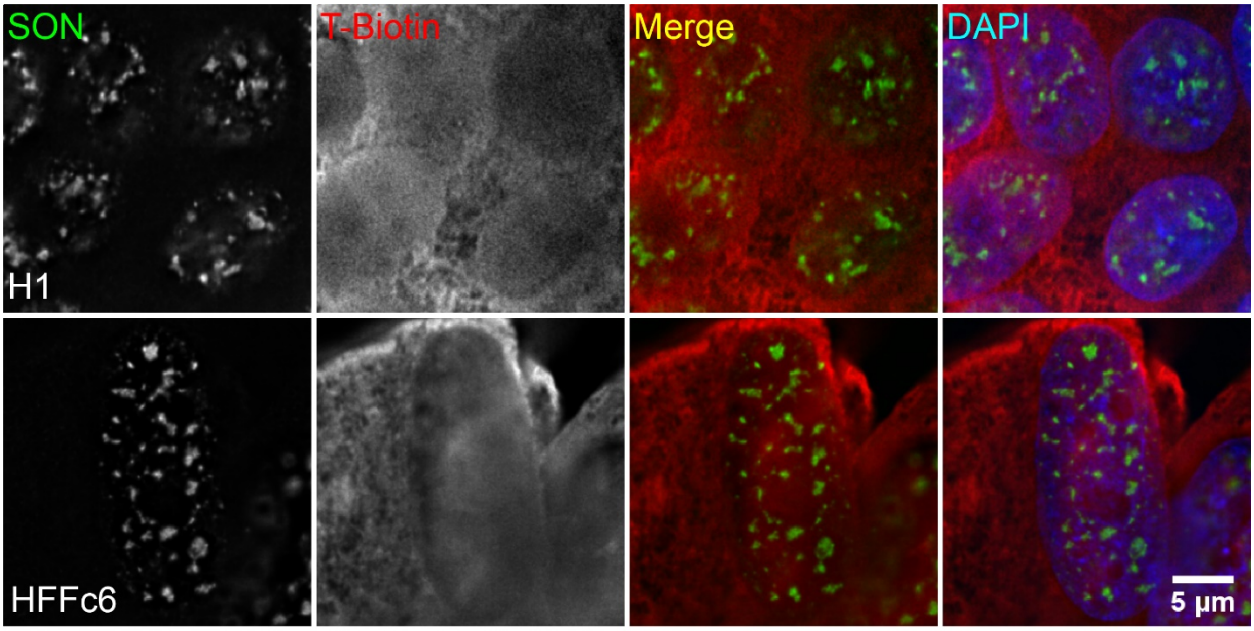


d

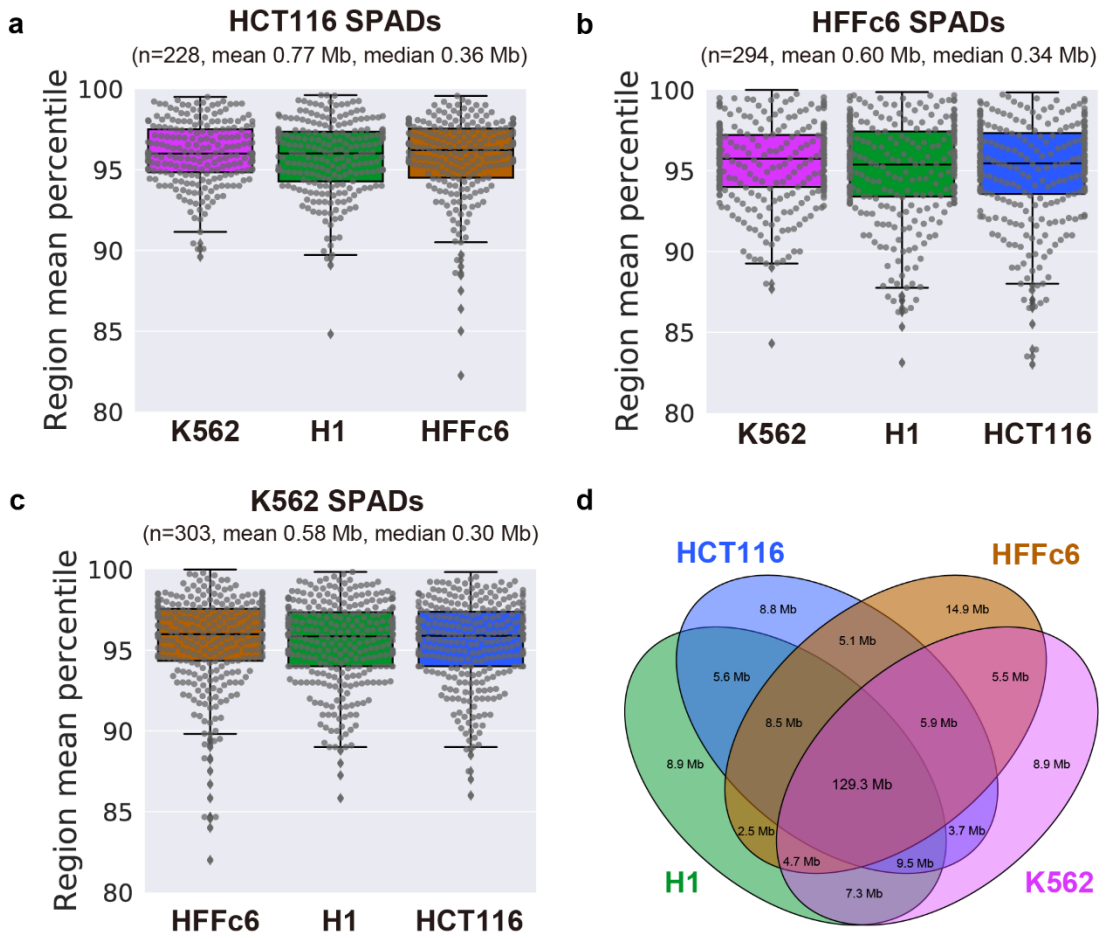
Cond.	Cell N (millions)	Pulldown DNA (ng)
A	440	4.8
C	220	9.3
E	220	59.3

Supplementary Figure 5. 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

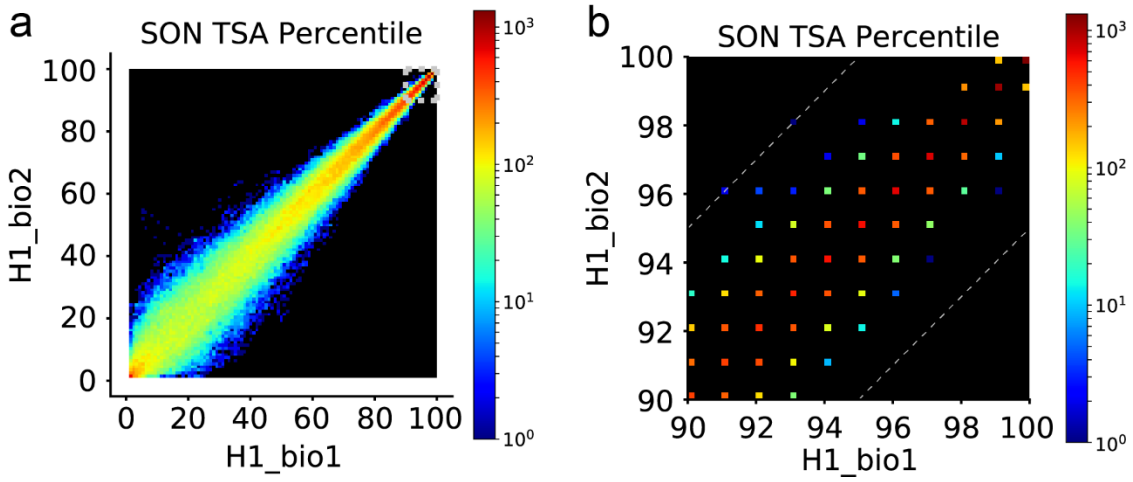
intensities along line profiles spanning nuclear speckles in **a)** for Conditions A, C, and E; **c)** Similar SON TSA-Seq enrichment score profiles (chromosome 3, 20 kb bins) for Conditions A, C, and E; **d)** Cell numbers used and pulldown DNA yields for Conditions A, C, and E.



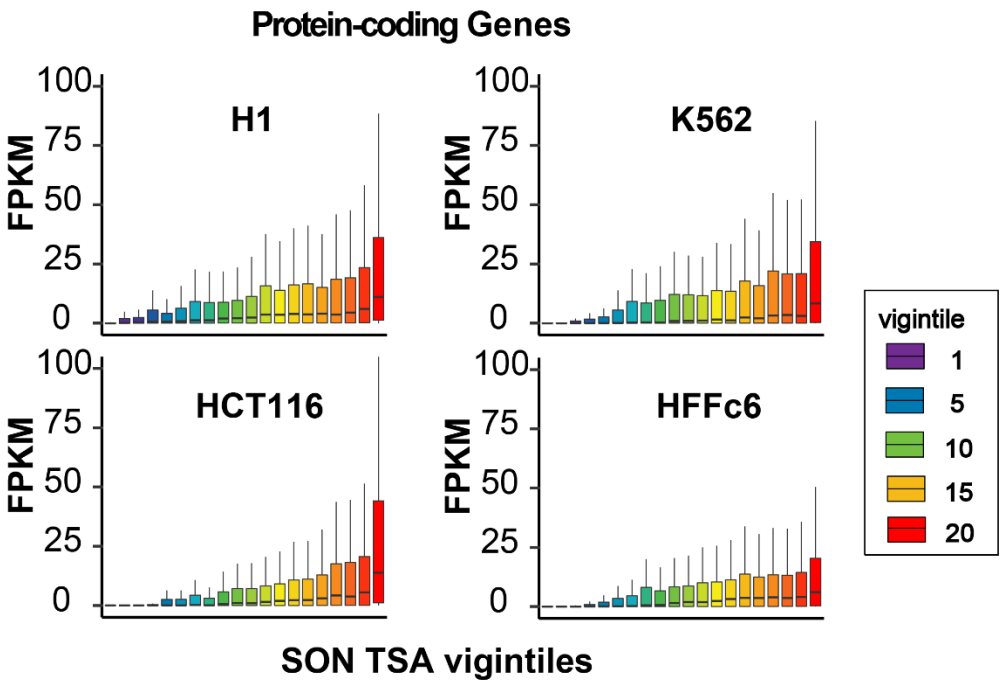
Supplementary Figure 6. Microscopy assays of biotin-labeling after Condition E SON TSA-labeling in H1 and HFFc6. Cells were immunostained to verify biotin-labeling after Condition E SON TSA-Seq: (Left to right) SON immunostaining, streptavidin staining of tyramide-biotin, merged image (SON, green; biotin, red), merged image plus DAPI (blue) in H1 (top) and HFFc6 (bottom) cells.



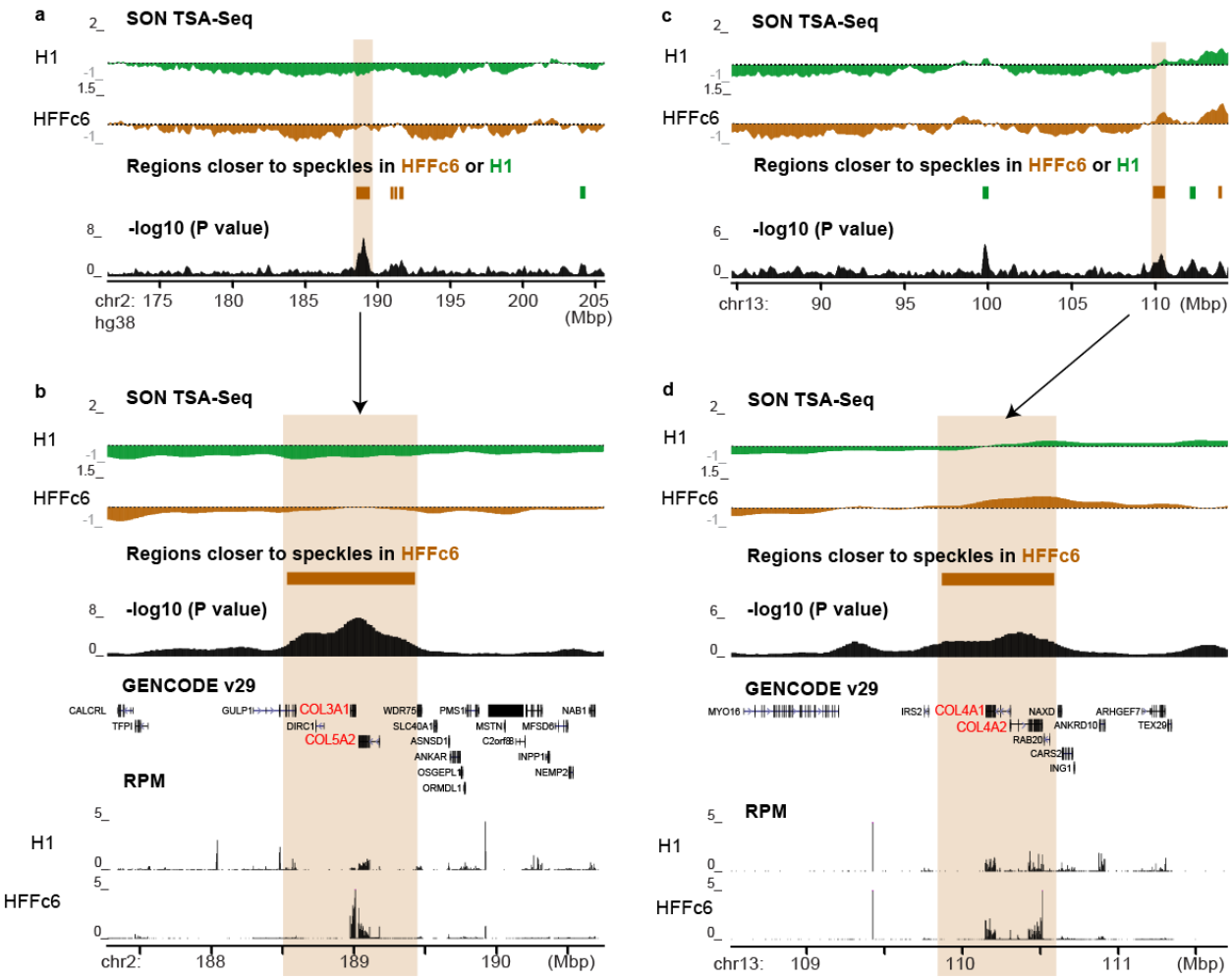
Supplementary Figure 7. Comparison of SPADs from HCT116, HFFc6, and K562 cells with SPADs in other cell lines. TSA-Seq percentile score distributions of SPADs from HCT116 **a)**, HFFc6 **b)**, and K562 **c)** in the other 3 cell lines. Box plot displays are as described in Fig. 2b legend. **d)** 4-way 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.



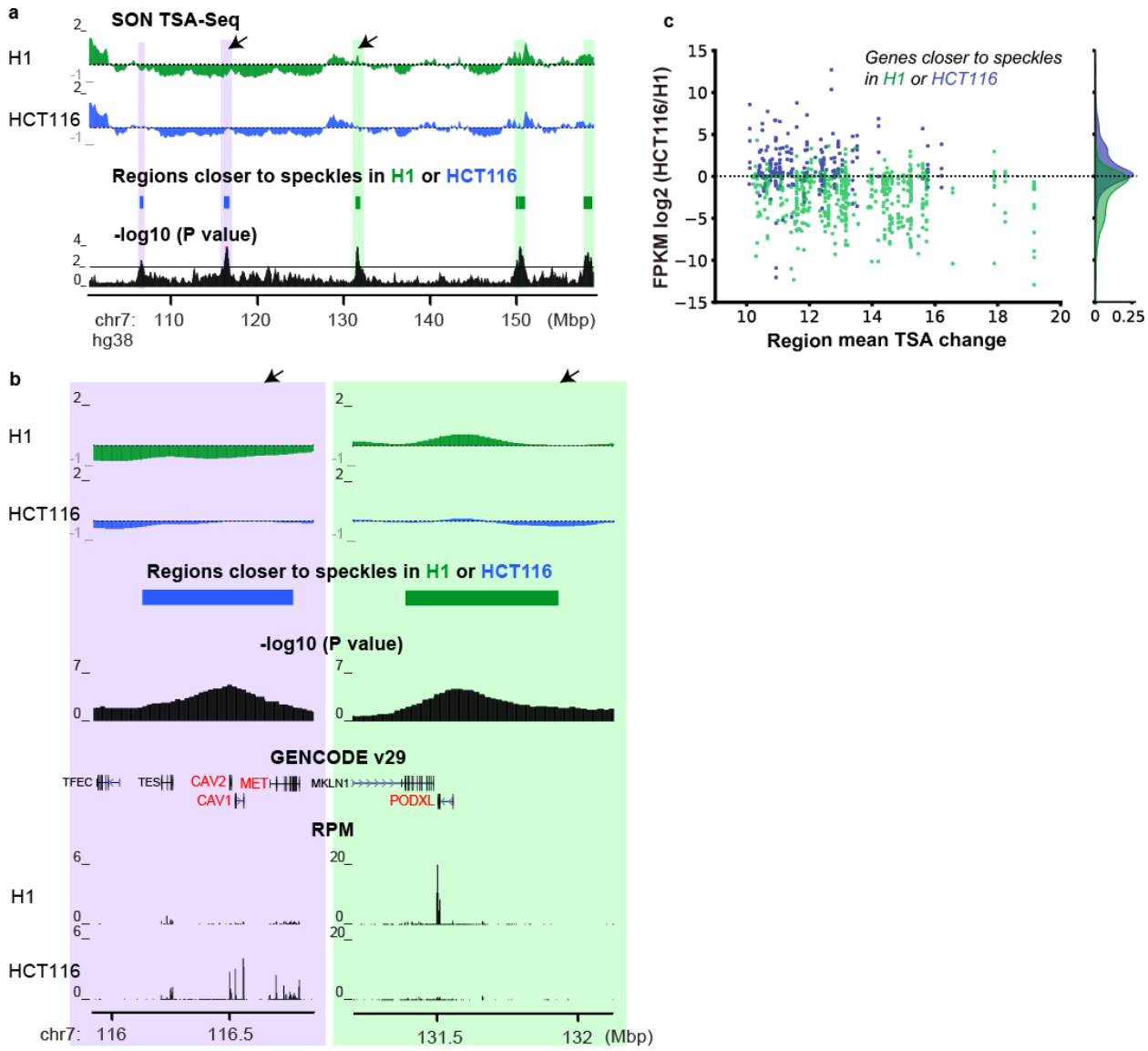
Supplementary Figure 8. Correlation between biological replicates of H1 SON TSA-Seq percentile scores. 2D histograms show correlation of SON TSA-Seq percentile scores (20 kb bins) for biological replicates 1 (x-axis) and 2 (y-axis) across entire percentile range (**a**) and top 10 percentile regions (**b**); **a-b)** Colors represent numbers of 20 kb bins with given TSA-Seq percentile scores in the 2 replicates falling within given replicate 1 percentile: replicate 2 percentile histogram 2D bin (1 percentile x 1 percentile intervals). **b)** Regions between dashed lines show histogram bins in which TSA-Seq percentile scores are within a 5 percentile difference of each other in the two replicates.



Supplementary Figure 9. Protein-coding gene expression increases with higher SON TSA-Seq enrichment score vigintiles in all four cell lines. Protein-coding gene FPKM in 20 SON TSA-Seq enrichment score vigintiles (division of percentile scores, 0-100, 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).

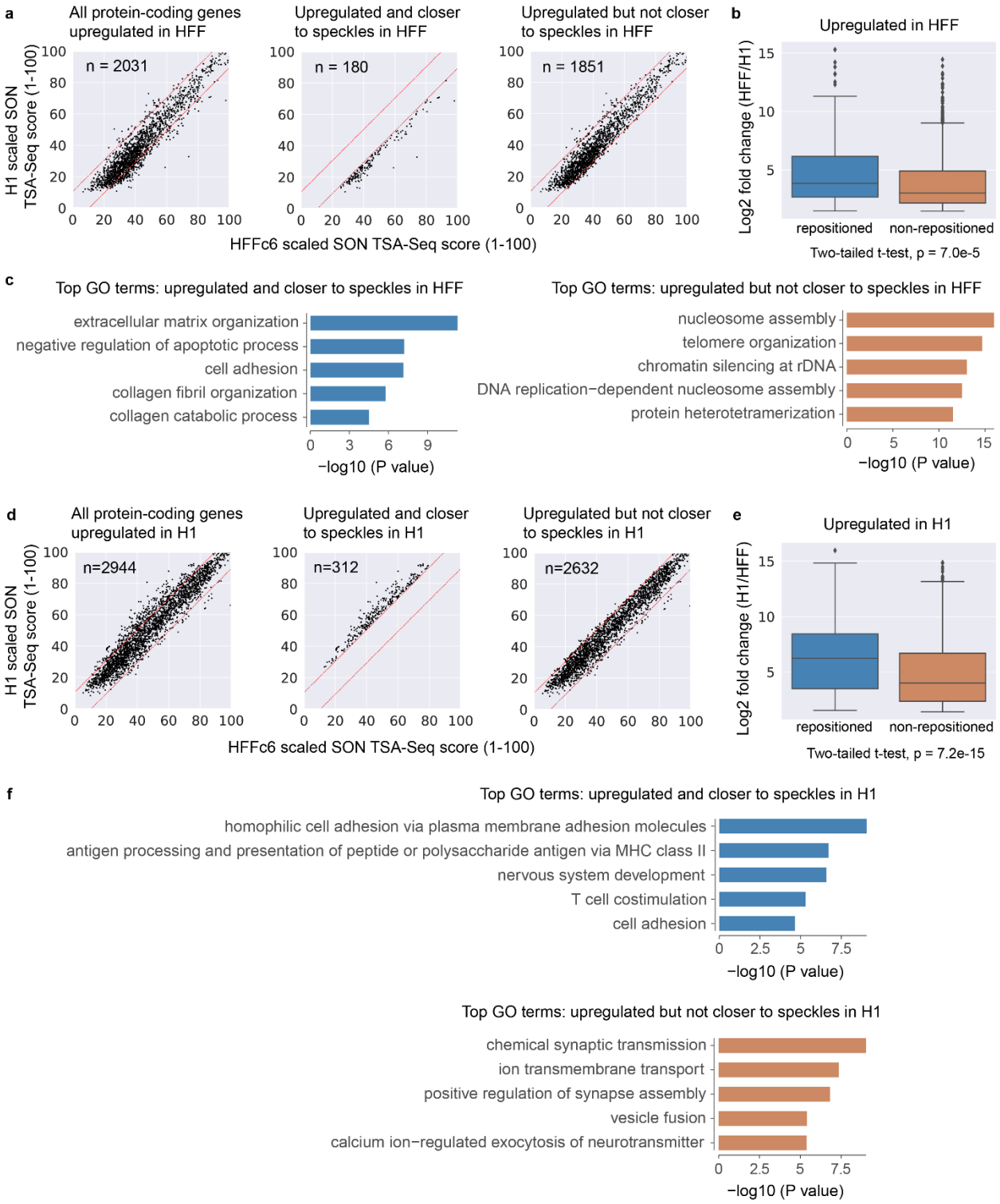


Supplementary Figure 10. Increased Collagen gene expression in HFF fibroblasts versus H1 hESCs correlates with closer relative distance to nuclear speckles in fibroblasts versus hESCs. **a, c)** Comparison of H1 and HFFc6 smoothed SON TSA-Seq enrichment scores: top track- H1 (green), middle track- HFFc6 (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.



Supplementary Figure 11. Strong bias towards increased (decreased) gene expression with repositioning of chromosome regions closer to (further from) nuclear speckles in H1 versus HCT116 cells. a) H1 and HCT116 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 - log10 (p-value per bin for significance of change in scaled SON TSA-Seq scores) (bottom); **b)** Zoomed view of two regions (arrows in **a**)

plus gene annotation (GENCODE v29) and RNA-Seq RPM values; **c)** Changes in expression of protein-coding genes in repositioned domains: scatterplots show log2 fold-changes in FPKM ratios (y-axis) between HCT116 and H1 versus absolute values of changes in mean scaled TSA-Seq scores (x-axis, max-min normalized: 1-100). Green (blue) dots are genes closer to speckles in H1 (HCT116). Gene distribution is shown as Kernel density plots on the right.



Supplementary Figure 12. Upregulated genes in regions that position relatively closer to nuclear speckles show modest skewing towards higher gene expression but significant

enrichment in cell-type specific gene-ontology classes. a) Scatter plots show scaled SON TSA-Seq scores (max-min normalized: 1-100) in H1 hESCs (y-axis) versus HFFc6 (x-axis) for protein-coding genes (dots) that are upregulated in HFF vs H1 for: left panel- all upregulated genes, middle panel- upregulated genes in regions relatively closer to nuclear speckles in HFFc6 cells, right panel- upregulated genes in regions that are not closer to nuclear speckles in HFFc6 cells. Red dashed lines represent thresholds for domains that show statistically significant changes in scaled HFFc6 vs H1 SON TSA-Seq scores. **b)** Comparison of log2-fold changes in gene expression in HFF versus H1 hESCs for protein-coding genes significantly upregulated in HFF cells for: upregulated genes in regions that are relatively closer to nuclear speckles in HFFc6 cells (left, blue) versus upregulated genes in regions that are not relatively closer to nuclear speckles in HFFc6 cells (right, orange). 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). **c)** Top gene ontology (GO) terms for protein-coding genes upregulated in HFF versus H1 cells for: left panel, blue- upregulated genes in regions that are relatively closer to nuclear speckles in HFFc6 versus H1 cells, right panel, orange- upregulated genes in regions that are not relatively closer to nuclear speckles in HFFc6 versus H1 cells. **d)** Scatter plots show scaled SON TSA-Seq scores (max-min normalized: 1-100) in H1 hESCs (y-axis) versus HFFc6 (x-axis) for protein-coding genes (dots) that are upregulated in H1 vs HFF for: left panel- all upregulated genes, middle panel- upregulated genes in regions relatively closer to nuclear speckles in H1 cells, right panel- upregulated genes in regions that are not closer to nuclear speckles in H1 cells. Red dashed lines represent thresholds for domains that show statistically significant changes in scaled H1 vs

HFFc6 SON TSA-Seq scores. **e)** Comparison of log2-fold changes in gene expression in H1 versus HFF cells for protein-coding genes significantly upregulated in H1 ESCs for: upregulated genes in regions that are relatively closer to nuclear speckles in H1 ESCs (left, blue) versus upregulated genes in regions that are not relatively closer to nuclear speckles in H1 ESCs (right, orange). 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). **f)** Top gene ontology (GO) terms for protein-coding genes upregulated in H1 versus HFF cells for: left panel, blue- upregulated genes in regions that are relatively closer to nuclear speckles in H1 versus HFFc6 cells, right panel, orange- upregulated genes in regions that are not relatively closer to nuclear speckles in H1 versus HFFc6 cells.

Supplementary Table 1

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

Supplementary Table 2

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

Supplementary Table 3

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

Supplementary Table 4 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.

Supplementary Table 5 is a separate Excel file showing results of RNA-Seq data analysis and correlation TSA-Seq data in the four cell lines.

Supplementary Table 6 is a separate Excel file showing all identified differentially expressed protein-coding genes between H1 and HFF cells.

Supplementary Table 7 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.