Supplemental material

Chen et al., https://doi.org/10.1083/jcb.201807108

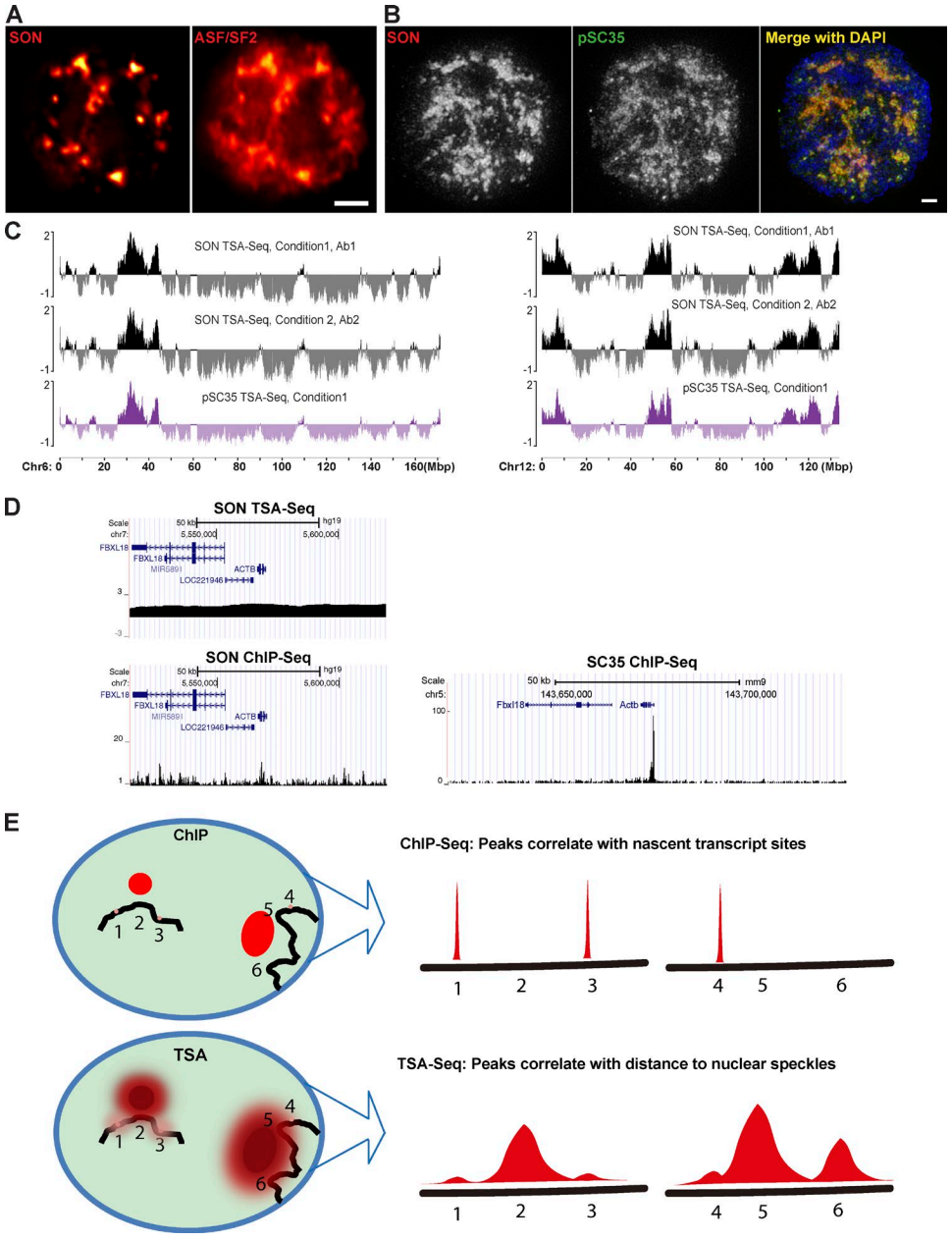
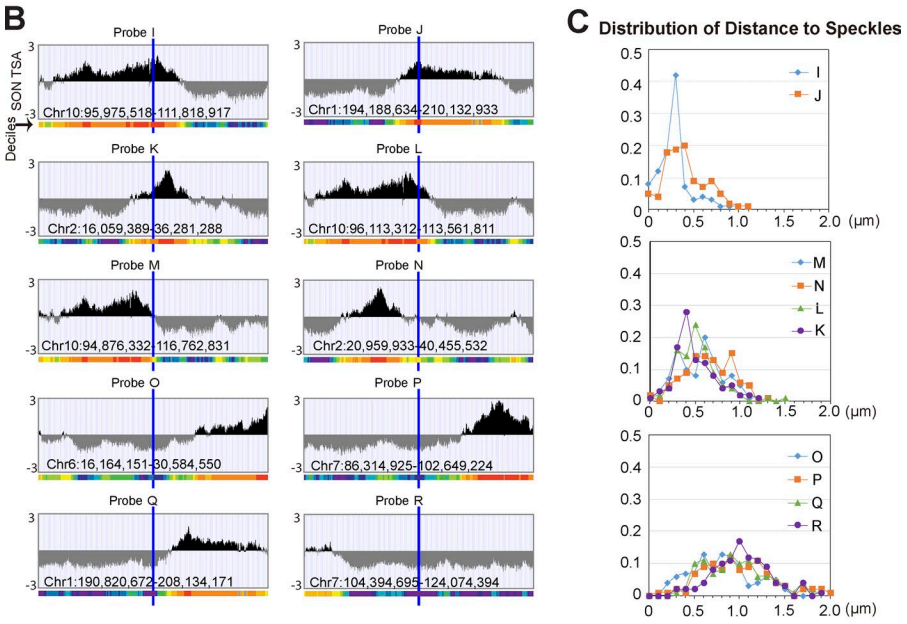


Figure S1. **TSA-Seq measures cytological proximity. (A)** SON immunostaining (left) provides much higher speckle to nucleoplasmic ratio than common speckle markers such as SR protein ASF/SF2 (right) as revealed by color heatmap display of widefield light microscopy–deconvolved optical sections. Bar, 2 µm. **(B)** Left to right: 3D SIM imaging of SON, pSC35, and merged image with SON (red), pSC35 (green), and DAPI (blue). Images are projections from 15 z sections with a 0.125-µm focus step size. Bar, 1 µm. **(C)** Speckle TSA-Seq genomic plots generated with two different SON antibodies or pSC35 antibody for chromosome 6 (left) and chromosome 12 (right) all show near-identical results: SON Ab1 (top) and Ab2 (middle) or pSC35 antibody (bottom). **(D)** Genomic plots comparing SON (Kim et al., 2016) or pSC35 (Ji et al., 2013) ChIP-Seq and TSA-Seq signals near the ∼100-kb region flanking the *ACTB* gene in human (left, SON) or mouse (right, pSC35). **(E)** TSA-Seq is fundamentally different from ChIP-Seq: The TSA signal, generated by staining a protein concentrated in speckles (red ovals) spreads over micrometer distances (left), thereby generating a broad genomic signal (right), while ChIP requires molecular contact of protein with DNA, thereby producing a localized signal (right) only from protein present over specific DNA sequences (left, small red dots, sites 1, 3, and 4). The speckle-derived TSA signal will dominate the overall TSA-Seq signal in proportion to the relative concentrations of the tyramide free radicals diffusing from the high concentration of speckle-localized protein versus from the small amount of localized SON protein near the DNA as visualized directly by light microscopy. Thus, the SON TSA-Seq signal spans large genomic regions (D, top left), showing near constant levels over tens of kilobase regions.

A

Probe	Genome Coordinates	SON-TSA log2 value	SON-TSA Decile Group	Average Distance to Speckles(μm)
A	Chr17:79,796,814-79,969,288	2.93	10	0.09
B	Chr6:31,694,290-31,908,943	2.76	10	0.11
C	Chr7:100,068,334-100,262,958	2.54	10	0.16
D	Chr18:46,328,263-46,524,378	0.53	9	0.47
E	Chr2:24,998,185-25,209,742	0.27	8 (9)	0.5
F	Chr17:69,698,111-69,877,392	-0.99	(3) 4	0.81
G	Chr2:22,925,891-23,120,000	-1.29	1	0.97
H	Chr18:38,612,910-38,817,071	-1.54	1	0.98
I	Chr10:103,818,001-103,976,434	1.83	10	0.25
J	Chr1:202,081,063-202,240,505	1.53	10	0.36
K	Chr2:26,069,230-26,271,448	0.73	9	0.46
L	Chr10:104,750,320-104,924,804	0.68	9	0.47
M	Chr10:105,710,150-105,929,014	-0.14	7 (8)	0.51
N	Chr2:30,610,255-30,805,210	-0.17	7	0.60
O	Chr6:23,302,249-23,446,452	-1.05	3 (2)	0.76
P	Chr7:94,400,404-94,563,746	-1.16	2	0.93
Q	Chr1:199,390,855-199,563,989	-1.32	1	0.92
R	Chr7:114,136,147-114,332,943	-1.37	1	0.97



D

	Fit from Microscopy			Fit from TSA-Seq		
	Sample 1	Sample 2	Sample 3	Replicate 1	Replicate 2	Replicate 3
Condition 1	-3.17±0.16	-3.16±0.17	-3.23±0.28	-3.44±0.56	-3.58±0.56	-3.46±0.69
Condition 2	-4.35±0.35	-4.24±0.26	-5.25±0.56	-4.75±0.34	-4.64±1.07	-6.00±1.35
Condition 3	-8.31±0.94	-8.26±0.86	-9.67±1.46	-8.95±1.31	N/A	N/A

Figure S2. **3D immuno-FISH validation that SON TSA-Seq score measures cytological mean distance to nuclear speckles. (A)** Genome location of 18 BAC probes (A–R) chosen from regions of different SON TSA-Seq scores and deciles. SON TSA-Seq scores were averaged over the ∼200-kb BAC FISH probe regions; the TSA decile group shown corresponds with the majority of the probe region, while the value in parentheses shows decile for smaller part of BAC probe. Average distance is the mean of distance distributions measured by 3D immuno-FISH. **(B)** SON TSA-Seq genomic plots flanking the additional 10 testing probe locations (blue bars). **(C)** Distance distribution to the nearest nuclear speckle for each of the 10 additional probes as measured by 3D immuno-FISH (n = 100). **(D)** Decay constant (R_0) and SEM for exponential fitting of TSA microscopy (as in Fig. 1 B) versus SON TSA-Seq genomic (as in Fig. 3 A) data for the three TSA staining conditions. The exponential decay constants from the direct cytological measurement of TSA spreading from lac repressor–stained lac operator arrays are within experimental error of the decay constants measured from the fit to the TSA-Seq genomic data for each of the three TSA staining conditions.

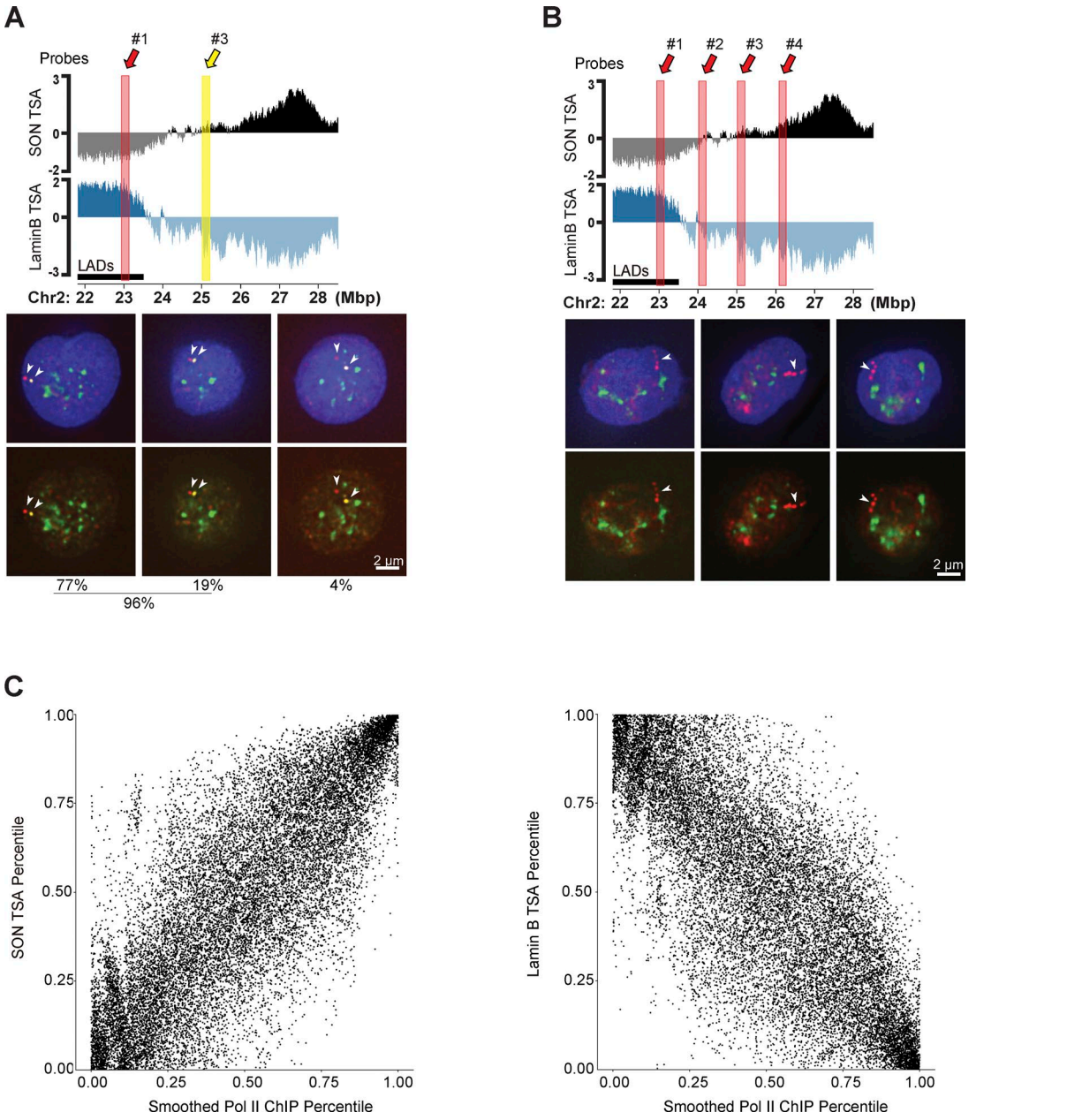


Figure S3. **TSA-Seq maps predict both chromosome location and chromosome topology. (A and B)** Top: SON and lamin B TSA-Seq genomic plots showing locations and colors of probes used to map trajectory orientation and path. Bottom: SON speckle staining (green), probe positions (arrowheads), and DAPI (blue) DNA staining are completely (top) or partially (bottom; no DAPI) merged. **(A)** 77% of examples show probe 1 maps ≤0.5 µm from the periphery (in 2D) with probe 2 located more interiorly (left); 19% of examples show probe 1 closer to the nuclear periphery than probe 2 but located >0.5 µm from the periphery (middle); 4% of examples show an indeterminate orientation (right; *n* = 114). **(B)** Chromosome trajectories point from nuclear periphery toward nuclear speckle in 69% of cases (85/124). **(C)** In contrast with SON versus lamin TSA-Seq (Fig. 5 B), Pol II ChIP signal poorly correlates with TSA-Seq. 2D scatter plots show genome-wide relationship between SON or lamin B TSA-Seq and Pol II ChIP-Seq. Each dot represents a single 160-kb bin plotted at a position corresponding with the average TSA-Seq percentiles and 1 Mbp smoothed normalized RNA Pol II ChIP-Seq percentiles.

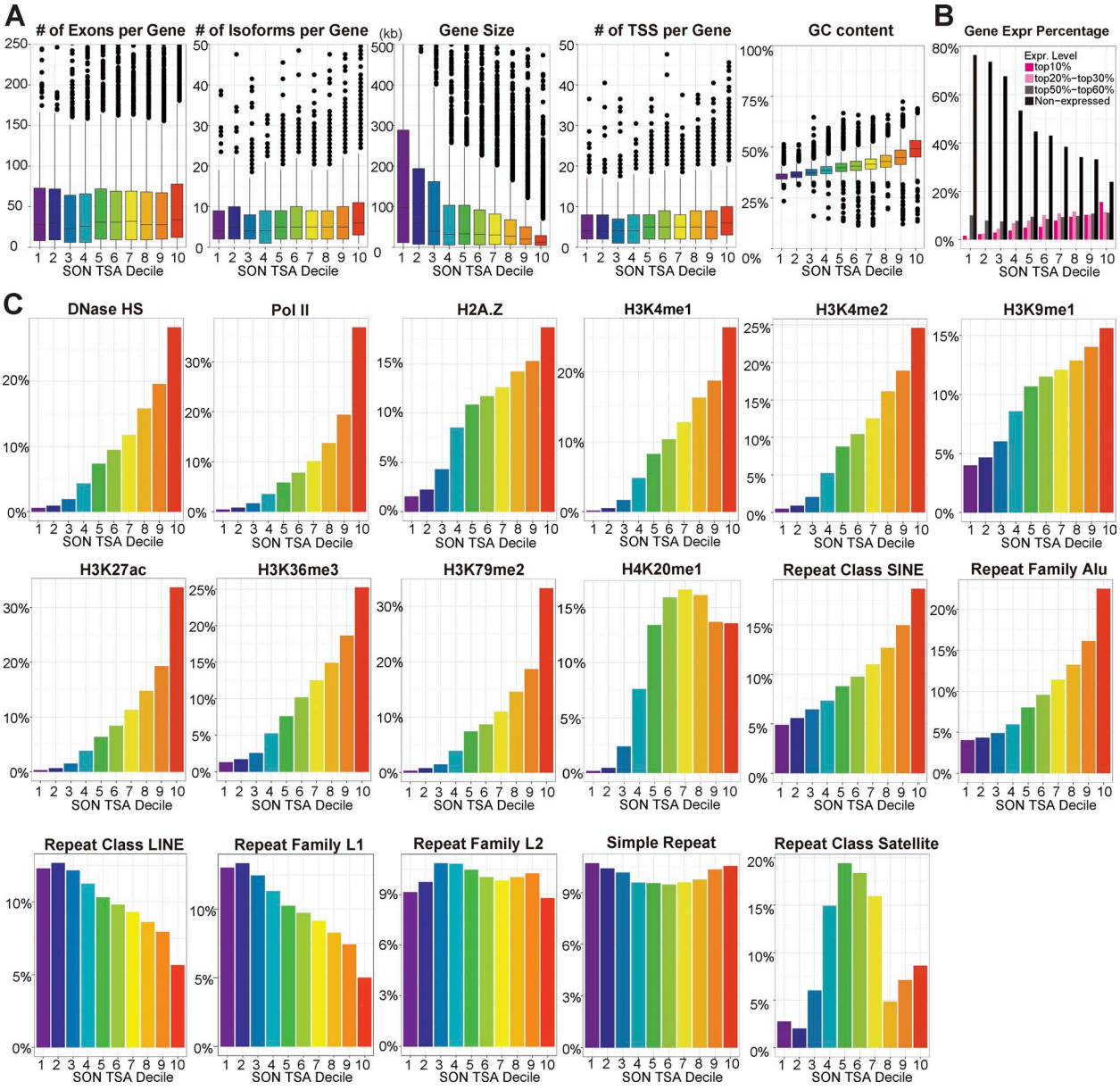


Figure S4. **Genome features relative to SON TSA-Seq decile. (A)** Box plots show genomic features in each of the SON TSA-Seq deciles. With increasing deciles, exon number and number of isoforms per gene show little change, gene size decreases, transcription start sites per gene increase slightly, and GC content increases progressively. **(B)** Percentage of highly expressed genes (top 10%; dark pink) increases, while percentage of nonexpressed genes (black) decreases with increasing decile. Genes expressed at near average levels (50–60%; gray) make up a near constant fraction of expressed genes in each decile, while genes expressed in the top 20–30% percentile (light pink) are relatively depleted from the bottom three deciles but otherwise show little variation. **(C)** Chromatin features related to transcription increase in a continuous gradient with increasing decile: DNase I hypersensitive sites peak count, RNA Pol II level, H2A.Z level, H3K4me1 peak count, H3K4me2 peak count, H3K9me1 level, H3K27ac level, H3K36me3 level, and H3K79me2 peak count. A roughly binary division is seen for H4K20me1 level in the 10 deciles. An increasing linear gradient relative to speckle proximity is observed for the SINE repeat class and Alu repeat family. A decreasing linear gradient is observed for the LINE repeat class and L1 repeat family. Instead, the L2 repeat families and the simple repeat class are uniformly distributed. Satellite repeats show a strong enrichment in the middle deciles. **(D)** Distribution of each Hi-C subcompartment type across SON TSA-Seq deciles.

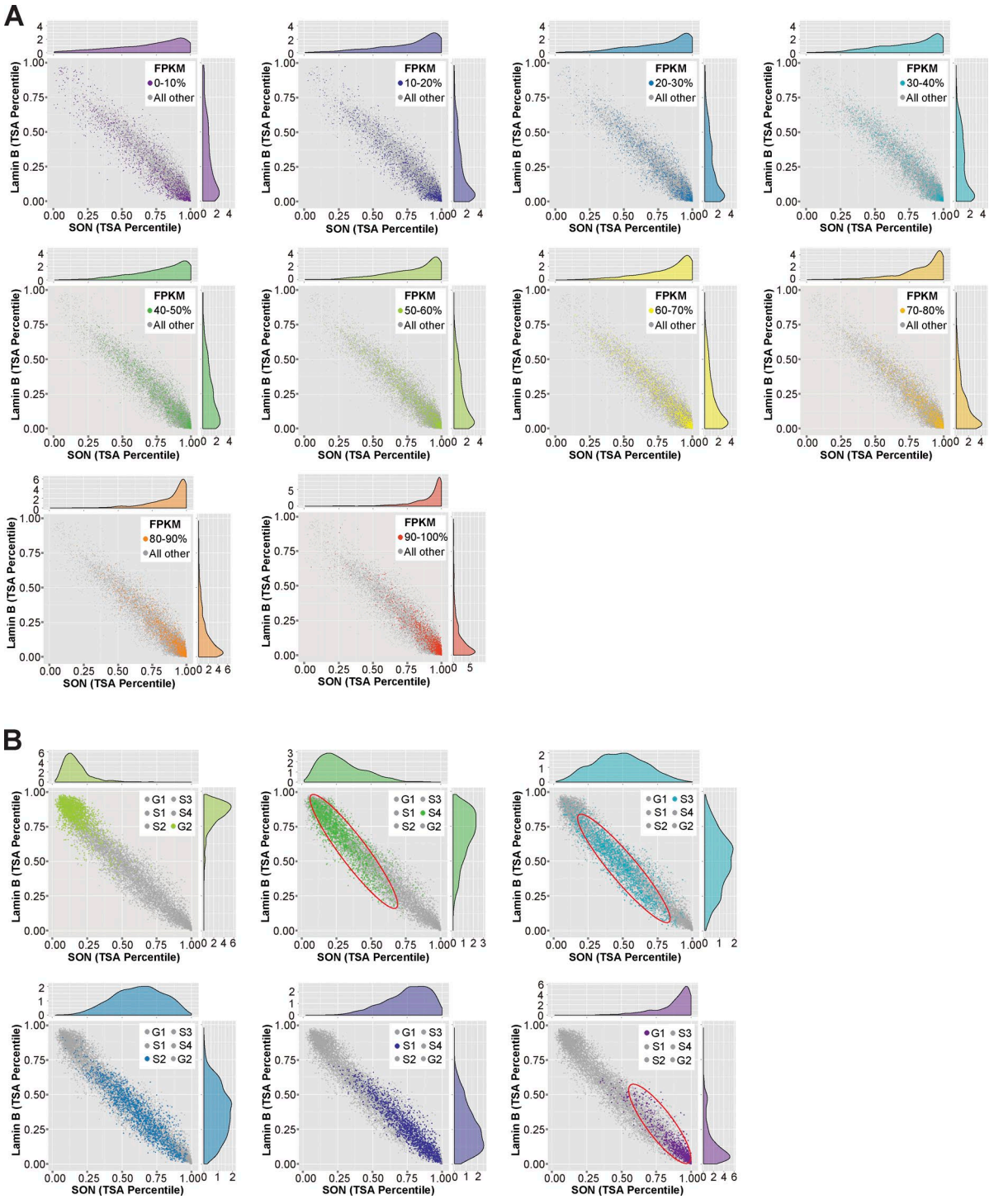


Figure S5. **2D scatter plots showing distribution of gene expression and DNA replication timing relative to speckle–lamin TSA-Seq axis. (A)** Protein-coding expressed genes are divided into 10 deciles according to their expression level. Each gene expression decile is plotted in its own 2D lamin B versus SON TSA-Seq percentile scatter plot, with genes in this decile appearing as colored dots; all other expressed genes appear as gray dots. Projections are shown on top and right sides. Increasing expression deciles are plotted left to right and top to bottom. As expression levels increase, gene location is skewed increasingly closer to speckles (x axis; higher SON TSA-Seq percentiles) and further from the lamina (y axis; lower lamin B TSA-Seq percentiles). With the highest expression decile, there is a skewing or bias toward disproportionally higher SON TSA-Seq percentiles for regions far from the nuclear lamina and close to speckles (80–100% SON TSA-Seq percentile; 0–20% lamin B TSA-Seq percentile). **(B)** Six DNA replication timing groups are plotted separately in 2D lamin B versus SON TSA-Seq percentile scatter plots. Each dot represents a 300-kb bin. DNA replication timing shifts progressively along the speckle–lamina axis with late replicating regions concentrated in areas near the lamina but far from speckles, and early replicating regions concentrated in areas near speckles but far from the lamina. Red ellipses show the slight skewing of earlier replicating regions toward disproportionally higher SON TSA-Seq percentiles and intermediate replicating regions toward lower lamin B TSA-Seq percentiles.

Table S1. **Summary of conditions for all TSA-Seq experiments**

Experiment number	Sample	Used for figures	Number of cells	1° Ab conc.	2° Ab conc.	TSA condition	Round of purification	Pulldown amount	Round of PCR amplification
1	No primary control for SON TSA-Seq	Yes	268 million	N/A	1:1,000	TSA	1	2.87 ng	15
1	SON TSA-Seq Ab1 condition 1	Yes	268 million	1:2,000	1:1,000	TSA	1	35.26 ng	15
1	SON TSA-Seq Ab1; condition 1 s pulldown	Yes					Started with 27.36 ng input from the first pulldown	19.92 ng	15
1	SON TSA-Seq Ab1; condition 2	Yes	268 million	1:2,000	1:1,000	TSA +Sucrose	2	7.888 ng	15
2	SON TSA-Seq Ab1; condition 1 Replicate	No	415 million	1:1,000	1:1,000	TSA	2	5.88 ng	8
2	SON TSA-Seq Ab1; condition 2 Replicate	No	415 million	1:1,000	1:1,000	TSA + sucrose	2	2.492 ng	8
2	SON TSA-Seq Ab1; condition 3	Yes	590 million	1:1,000	1:1,000	TSA + sucrose + DTT	2	~1-1.5 ng	8 for input; 12 for pulldown
3	SON TSA-Seq Ab2; condition 2	Yes	270 million	1:2,000	1:1,000	TSA + sucrose	1	10.504	8
4	No primary control for lamin A/C TSA-Seq	No	150 million	N/A	1:200	TSA	2	<1.122 ng	8 for input; 10 for pulldown
4	Lamin A/C TSA-Seq; condition 2	No	150 million	1:1,000	1:200	TSA + sucrose	2	3.78 ng	8
5	No primary control for lamin A/C TSA-Seq replicate	No	710 million	1:1,000	1:1,000	TSA + sucrose	1	1.965 ng	12 for input; 16 for pulldown
5	Lamin A/C TSA-Seq; condition 2 replicate	Yes	710 million	1:1,000	1:1,000	TSA + sucrose	1	23 ng	12
6	No primary control for lamin B TSA-Seq	No	400 million	N/A	1:1,000	TSA + sucrose	1	1.364 ng	10
6	Lamin B TSA-Seq; condition 2	Yes	400 million	1:1000	1:1,000	TSA + sucrose	1	4.8279 ng	10
7	No primary control for lamin B TSA-Seq replicate	No	460 million	N/A	1:1,000	TSA + sucrose	1	3.9 ng	12
7	Lamin B TSA-Seq condition 2 replicate	No	460 million	1:1000	1:1,000	TSA + sucrose	1	6.85 ng	12
8	pSC35 TSA condition 1	No	485 million	1:500	1:1,000	TSA	1	32.136	8
9	No primary control for pSC35 TSA-Seq replicate	No	326 million	N/A	1:1,000	TSA	1	3.435 ng	12
9	pSC35 TSA-Seq condition 1 replicate	Yes	326 million	1:500	1:1,000	TSA	1	8.9 ng	12
10	Pol II TSA-Seq condition 1	No	225 million	1:300	1:1,000	TSA	1	52.44 ng	8

Table S1. **Summary of conditions for all TSA-Seq experiments *(Continued)***

Experiment number	Sample	Used for figures	Number of cells	1° Ab conc.	2° Ab conc.	TSA condition	Round of purification	Pulldown amount	Round of PCR amplification
11	No primary control for Pol II TSA-Seq replicate	No	175 million	N/A	1:1,000	TSA	1	2.3 ng	10 for input; 12 for pulldown
11	Pol II TSA-Seq condition 1 replicate	Yes	175 million	1:300	1:1,000	TSA	1	14.05 ng	10

Table S2. **BACs used for DNA FISH**

Probes	BAC	Genome coordinates
Probes used to generate exponential fitting		
A	RP11-634L10	Chr17:79,796,814-79,969,288
B	RP11-479I13	Chr6:31,694,290-31,908,943
C	RP11-264N5	Chr7:100,068,334-100,262,958
D	RP11-1058N17	Chr18:46,328,263-46,524,378
E	CTD-3106L12	Chr2:24,998,185-25,209,742
F	RP11-997B19	Chr17:69,698,111-69,877,392
G	RP11-978O5	Chr2:22,925,891-23,120,000
H	RP11-846O11	Chr18:38,612,910-38,817,071
Probes used to test the fitting		
I	RP11-302K17	Chr10:103,818,001-103,976,434
J	RP11-246J15	Chr1:202,081,063-202,240,505
K	RP11-791D16	Chr2:26,069,230-26,271,448
L	CTD-3244P16	Chr10:104,750,320-104,924,804
M	CTD-2503D10	Chr10:105,710,150-105,929,014
N	RP11-729K13	Chr2:30,610,255-30,805,210
O	RP11-53I19	Chr6:23,302,249-23,446,452
P	CTD-3008G3	Chr7:94,400,404-94,563,746
Q	RP11-543G21	Chr1:199,390,855-199,563,989
R	RP11-1047B3	Chr7:114,136,147-114,332,943
For visualization of chromosome trajectories		
1	RP11-978O5	Chr2:22,925,891-23,120,000
2	RP11-1080G13	Chr2:24,019,044-24,218,791
3	CTD-3106L12	Chr2:24,998,185-25,209,742
4	RP11-791D16	Chr2:26,069,230-26,271,448
5	RP11-281H16	Chr2:27,201,851-27,389,538
6	RP11-373D23	Chr2:28,560,953-28,718,617
7	RP11-729K13	Chr2:30,610,255-30,805,210
8	RP11-1008C9	Chr2:34,145,606-34,336,642
9	RP11-302K17	Chr10:103,818,001-103,976,434
10	CTD-3244P16	Chr10:104,750,320-104,924,804
11	CTD-2503D10	Chr10:105,710,150-105,929,014
12	RP11-833F23	Chr10:106,768,289-106,971,464

Table S3 is a separate Excel file showing a list of protein-coding genes and their SON TSA-Seq deciles and expression levels.

Table S4 is a separate Excel file showing public datasets analyzed.

References

Ji, X., Y. Zhou, S. Pandit, J. Huang, H. Li, C.Y. Lin, R. Xiao, C.B. Burge, and X.D. Fu. 2013. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell.* 153:855–868. https://doi.org/10.1016/j.cell.2013.04.028

Kim, J.H., M.C. Baddoo, E.Y. Park, J.K. Stone, H. Park, T.W. Butler, G. Huang, X. Yan, F. Pauli-Behn, R.M. Myers, et al. 2016. SON and Its Alternatively Spliced Isoforms Control MLL Complex-Mediated H3K4me3 and Transcription of Leukemia-Associated Genes. *Mol. Cell.* 61:859–873. https://doi.org/10.1016/j.molcel.2016.02.024