Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler

Yu Chen[1] ⓘ, Yang Zhang[2,3], Yuchuan Wang[3] ⓘ, Liguo Zhang[1], Eva K. Brinkman[4], Stephen A. Adam[5], Robert Goldman[5], Bas van Steensel[4], Jian Ma[3], and Andrew S. Belmont[1,6,7] ⓘ

While nuclear compartmentalization is an essential feature of three-dimensional genome organization, no genomic method exists for measuring chromosome distances to defined nuclear structures. In this study, we describe TSA-Seq, a new mapping method capable of providing a "cytological ruler" for estimating mean chromosomal distances from nuclear speckles genome-wide and for predicting several Mbp chromosome trajectories between nuclear compartments without sophisticated computational modeling. Ensemble-averaged results in K562 cells reveal a clear nuclear lamina to speckle axis correlated with a striking spatial gradient in genome activity. This gradient represents a convolution of multiple spatially separated nuclear domains including two types of transcription "hot zones." Transcription hot zones protruding furthest into the nuclear interior and positioning deterministically very close to nuclear speckles have higher numbers of total genes, the most highly expressed genes, housekeeping genes, genes with low transcriptional pausing, and super-enhancers. Our results demonstrate the capability of TSA-Seq for genome-wide mapping of nuclear structure and suggest a new model for spatial organization of transcription and gene expression.

Introduction

While the human genome has been sequenced, how this linear genome sequence folds in 3D within the nucleus remains largely unknown. New genomic methods such as Hi-C (Lieberman-Aiden et al., 2009; Rao et al., 2014) have generated increasing interest in how 3D chromosome folding may regulate genome functions during development or in health and disease. However, these 3C (chromosome conformation capture)-based methods do not directly report on chromosome positioning within nuclei.

What is needed is an ability to translate microscopic views of DNA position relative to nuclear compartments (such as the nuclear lamina, nucleolus, or nuclear speckles) into genome-wide maps that show how close loci are to a given compartment and how the chromosomal fiber traverses between compartments. For example, whether transcriptionally active chromosome regions are targeted reproducibly to particular nuclear compartments has been a long-standing question. Using DNA FISH, a population-based, statistical shift toward the nuclear center has been observed for a number of genes undergoing transcriptional activation (Takizawa et al., 2008), leading to the proposal of a gradient of increased transcriptional activity from the nuclear

periphery to center (Takizawa et al., 2008; Bickmore, 2013). However, the functional significance of this radial positioning has been difficult to rationalize given the large variability of gene positioning within individual nuclei (Takizawa et al., 2008; Kölbl et al., 2012). Alternatively, this stochastic radial positioning of genes could be the consequence of a more deterministic positioning of genes relative to a nuclear compartment or compartments that themselves show a stochastic radial positioning.

Nuclear speckles, excluded from the nuclear periphery and enriched toward the nuclear center (Carter et al., 1991), are an excellent candidate for such a nuclear compartment. Nuclear speckles were first visualized by transmission EM (TEM) as dense clusters of 20–25-nm-diameter RNP granules (Fakan and Puvion, 1980) termed interchromatin granule clusters, and they have alternatively been proposed to be storage sites for RNA-processing components (Spector and Lamond, 2011) or transcription hubs for a subset of active genes (Xing et al., 1995; Shopland et al., 2003; Hall et al., 2006). Microscopic studies have demonstrated the very close association with (Xing et al., 1995; Moen et al., 2004) or even movement to (Hu et al., 2009; Khanna

[1]Department of Cell and Developmental Biology, University of Illinois at Urbana-Champaign, Urbana, IL;   [2]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL;   [3]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA;   [4]Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, Netherlands;   [5]Department of Cell and Molecular Biology, Northwestern University Feinberg School of Medicine, Chicago, IL;   [6]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL;   [7]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL.

Correspondence to Andrew S. Belmont: asbel@illinois.edu.

et al., 2014) nuclear speckles of a small number of genes upon transcriptional activation. One major problem, however, in judging the significance of this speckle association has been the absence of any successful genome-wide survey of the prevalence of gene association with nuclear speckles. The pooled results from several previous low-throughput microscopy surveys showed that approximately half of the 25 active genes examined had a "close" association to nuclear speckles (Hall et al., 2006), but this small sampling of active genes might not be representative of the entire genome. Another significant problem has been the nonquantitative assessment of "close" used in previous studies and the absence of any comparison to the percentage of the genome localized within similar distances.

Current genomic methods such as DNA adenine methyltransferase identification (DamID; Vogel et al., 2007) and chromatin immunoprecipitation (ChIP) sequencing (ChIP-Seq; Landt et al., 2012) are limited in mapping nuclear speckle–associated domains as they measure molecular contact frequencies with particular proteins but not the actual cytological distances from specific nuclear compartments. Nuclear speckles behave like a dynamic phase-separated body (Brangwynne, 2011; Zhu and Brangwynne, 2015; Galganski et al., 2017), and no detectable DNA is contained within them (Spector and Lamond, 2011), while serial-section TEM reconstructions showed large-scale chromatin fibers near, but not necessarily contacting, the periphery of these granule clusters (Belmont and Bruce, 1994). Thus, the first challenge is detecting DNA sequences that are close to nuclear speckles but not in direct contact. Moreover, all known speckle proteins, although enriched in nuclear speckles, are also found at lower concentrations throughout the nucleus; most of these proteins are involved in RNA processing and/or transcription and bound to chromatin or nascent transcripts. Therefore, the second challenge is producing a signal proportional to the total amount of a protein within a microscopic neighborhood rather than the small fraction of that protein in molecular contact with DNA.

Tyramide signal amplification (TSA), a well-developed technology widely used in immunocytochemistry, works by using HRP to generate tyramide free-radicals; these free radicals then diffuse and react to form covalent bonds with nearby macromolecules including the known target of protein tyrosine residues (Bobrow et al., 1989, 1991; van Gijlswijk et al., 1996, 1997). Two previous applications of TSA to genomics, RNA TRAP (tagging and recovery of associated proteins; Carter et al., 2002) and immuno-TRAP (Ching et al., 2013), used chromatin pulldown to assay TSA labeling. However, core histones are poor targets for TSA labeling, while neither of these previous applications of TSA adequately tested for possible labeling bias based on varying nonhistone protein identity and concentration along the genome.

In this study, we present TSA sequencing (TSA-Seq), the first genomic method capable of estimating cytological distances of chromosome loci genome-wide relative to a particular nuclear compartment and even inferring chromosome trajectories from one compartment to another. This distance mapping was enabled by our recognition that TSA labels DNA directly and uniformly across the genome and by our demonstration that the steady-state concentration of tyramide free radicals during the TSA reaction can be accurately modeled by a simple exponential decay.

TSA-Seq represents the first genomic method that starting from basic physical principles and equations converts sequencing reads into actual calibrated physical distances at a resolution superior to conventional light microscopy. Our TSA-Seq results in K562 cells suggest a refined model of nuclear organization in which the more biologically relevant metric is not fractional distance relative to the nuclear periphery versus nuclear center but rather is the distance of genomic regions to specific nuclear compartments, e.g., the nuclear lamina, speckles, or still unknown regions such as transcription factories.

Results

Gradient of diffusible free radical produced by TSA provides distance-dependent biotin labeling

To measure the relative distance of genomic regions to specific nuclear compartments, we turned to TSA (Wang et al., 1999), a commonly used immunochemistry method for amplifying signals for target detection. TSA uses an antibody-coupled HRP to catalyze the formation of diffusible biotin-tyramide free radicals and create a free radical concentration gradient centered at the staining target (Fig. 1 A).

We measured how far the TSA signal spreads using a cell line with multiple intranuclear GFP spots produced by EGFP–lac repressor binding to 256mer lac operator DNA tags contained within an integrated bacterial artificial chromosome (BAC) multicopy transgene array (Hu et al., 2009). Each GFP spot served as a point source for biotin-tyramide free radical generation after anti-GFP TSA. Diffusion of tyramide free radicals from a point source combined with a spatially invariant free radical degradation rate is predicted to produce an exponential decay in free radical concentration with distance from this point source at steady state (Wartlick et al., 2009). Using the line profiles of streptavidin staining of the observed biotin distribution, we fit the observed signals to exponential decay functions (Fig. 1 B). The unmodified TSA reaction product spread over an ~1-µm radius (condition 1). Adding sucrose to increase the solution viscosity reduced this spreading moderately (condition 2), although it reduced the biotin labeling approximately fourfold (not depicted). Adding both sucrose and DTT (condition 3) to reduce the tyramide free radical's lifetime and diffusion path length reduced biotin labeling ~20-fold (not depicted) but reduced the staining radius to less than ~0.5 µm. Here, the staining radius is defined as the radius at which the TSA labeling becomes indistinguishable from background staining.

To label nuclear speckles in K562 cells, we used anti-SON (Ab1; see Materials and methods) antibody staining against the protein SON, a highly specific marker for nuclear speckles (Fig. S1 A). Comparison of anti-SON immunostaining (Fig. 1 C, green) with the streptavidin labeling of the TSA tyramide-biotin signal (Fig. 1 C, red) shows TSA signal spreading over distances consistent with our estimates from the GFP-stained lac operator arrays (Fig. 1 B). The intranuclear biotin labeling is dominated by this TSA signal spreading from nuclear speckles, as shown by 3D surface plots (Fig. 1 D) for conditions 1–3, consistent with the low nucleoplasmic SON immunostaining (Figs. 1 C and S1 A). TSA reactions using no primary antibody showed minimal labeling (not depicted).
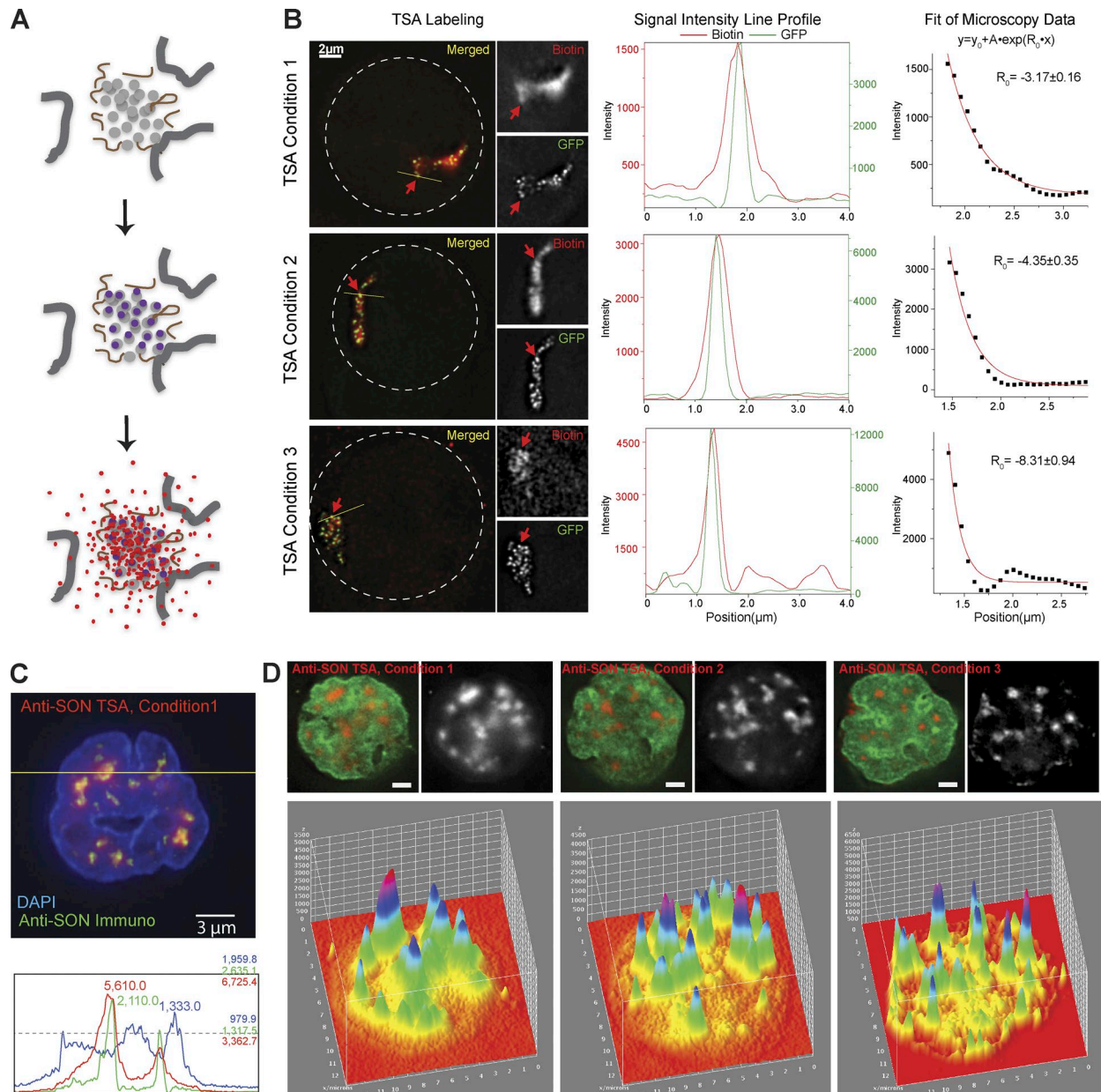
Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology   4026
https://doi.org/10.1083/jcb.201807108

Figure 1. **TSA generates biotin labeling, which decays exponentially from the staining target. (A)** TSA: Primary and HRP-conjugated secondary antibodies (purple dots) target epitopes (gray dots). HRP catalyzes formation of tyramide-biotin free radicals (red), which diffuse and covalently link to nearby proteins (gray dots), DNA (gray lines), and RNA (brown lines). **(B)** Exponential fitting of spreading of TSA staining: condition 1 (top), condition 2 (middle), and condition 3 (bottom). Left: Biotin staining (red) of TSA staining of GFP-lac repressor (green) bound to BAC transgene array; white dashes encircle nuclei and red arrows mark GFP spot used for line scan. Middle: GFP spot (green) versus biotin intensities (red) along yellow line scan (left). Right: Exponential fit of biotin intensity versus distance from GFP spot. **(C)** Image (top) and intensities (bottom) along yellow line scan (top) of SON immunostaining (green), biotin staining after TSA (red), and DNA DAPI staining (blue). **(D)** Images of SON TSA staining conditions 1–3. Top: Biotin staining (red) merged with DAPI DNA staining (green). Bottom: TSA staining surface plots. Bars, 2 µm.

TSA-Seq translates the distance-dependent TSA biotin labeling into a quantitative sequencing readout

To convert TSA staining into a genomic readout, we followed the TSA reaction in K562 cells with reversal of formaldehyde cross-linking, DNA isolation, pulldown of biotinylated DNA, and high-throughput sequencing (Fig. 2 A). Workflow quality control included microcopy of streptavidin-stained cell aliquots following the TSA reaction and anti-biotin dot-blot staining of the purified DNA (not depicted).

Two previous TSA applications focused on TSA protein labeling through the known reaction of tyramide free radicals with tyrosines and tryptophans (Carter et al., 2002; Ching et al., 2013). However, early in our exploration of TSA, we found that core histones showed minimal labeling by the biotin-tyramide free radical as determined by Western blotting of total biotinylated proteins after TSA staining (not depicted). This is likely due to the low reactivity of core histone tyrosines within the nucleosome; indeed, 13/15 of the native nucleosomal tyrosines are inaccessible to reac-
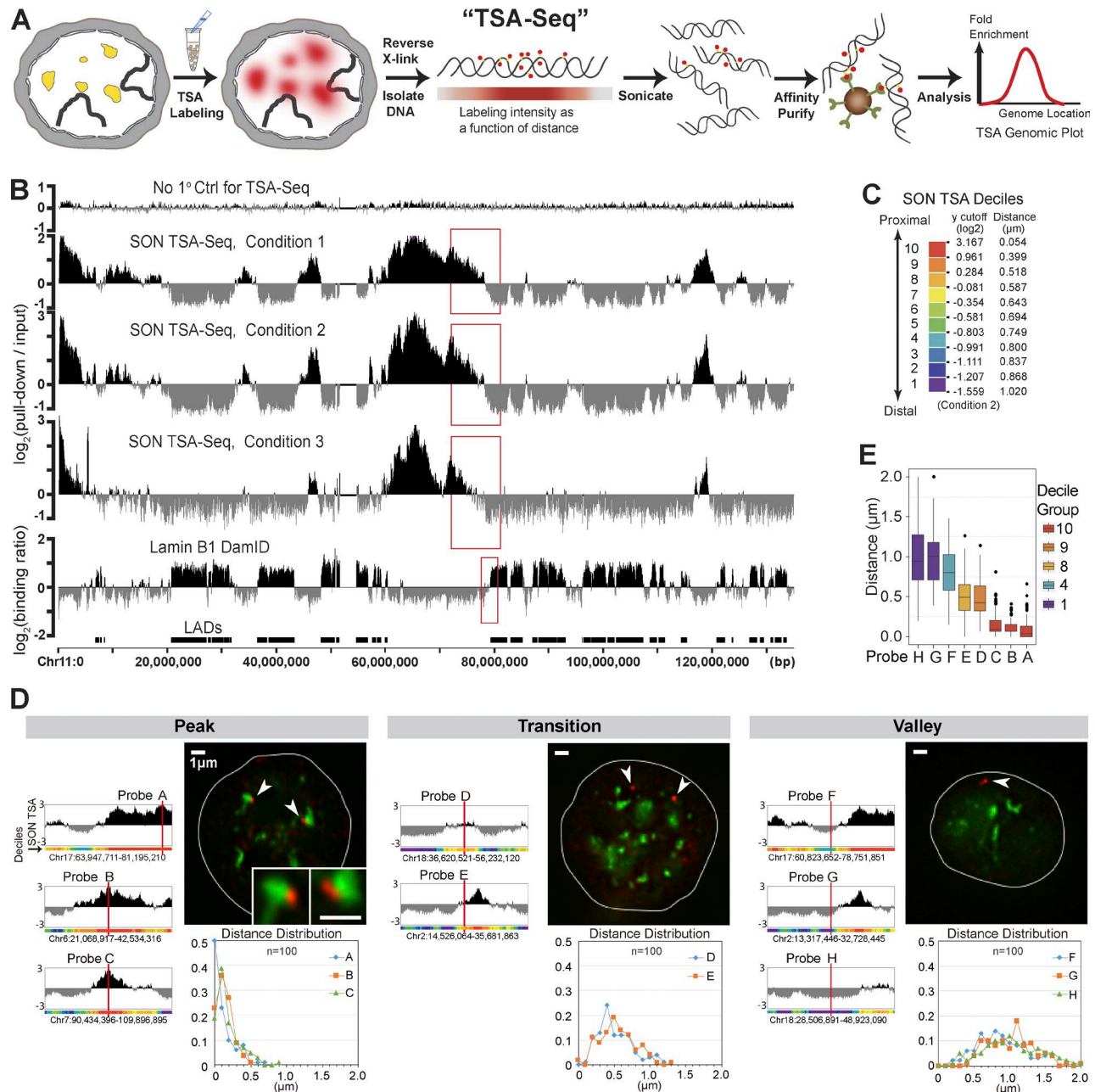
Figure 2. Genome-wide mapping chromosome organization relative to nuclear speckles using TSA-Seq. (A) Workflow schematic. **(B)** SON TSA-Seq genomic plots (human Chr11): no primary control (top), TSA staining conditions 1–3 (middle), and lamin B1 DamID (bottom). Red boxes indicate SON TSA-Seq peak-to-valley versus inter-LAD–LAD transitions. **(C)** Color-coding, TSA-Seq scores, and estimated speckle distances for 10 SON TSA-Seq deciles (TSA condition 2). **(D)** For each of the peak (probes A–C), transition (probes D and E), and valley (probes F–H) regions, SON TSA-Seq genomic plots flanking probe locations (red bars) are shown on left; histograms of probe distances to the nearest speckle (*n* = 100) are shown on right with representative 3D immuno-FISH images: SON (green), FISH (red), and nucleus (white lines). Bars, 1 µm. **(E)** Box plot of probe distances to nearest speckle (*n* = 100).

tion with p-nitrobenzenesulfonyl fluoride (Zweidler, 1992). The low TSA reactivity of core histones suggested the likelihood of a significant bias of chromatin TSA labeling toward regions with elevated nonhistone content, motivating us to examine whether DNA would react with the tyramide free radical. Indeed, TSA also labels DNA, as measured by dot-blot streptavidin staining of purified DNA isolated from cells stained by TSA or purified plasmid DNA reacted with free HRP and tyramide-biotin (not depicted). Using DNA labeling, we avoid potential biases introduced by differential chromatin solubilization and/or variations in the

tyrosine and tryptophan content of surrounding nonhistone proteins. As compared with ChIP-Seq, input variation across the genome is minimized, and pulldown specificity can be improved using harsher washing conditions.

We generated TSA enrichment genomic maps by plotting the log_2 ratio of the normalized pulldown read density versus the normalized read density of input DNA over a 20-kbp sliding window. We observed very similar enrichment maps for all three staining conditions (Fig. 2 B), with an ~16–32-fold enrichment/depletion dynamic range. Enrichment peaks were slightly

Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology   4028
https://doi.org/10.1083/jcb.201807108

sharper and the dynamic range higher for staining condition 2, which produced less spreading of the tyramide free radical. Condition 3 further sharpened larger TSA-Seq peaks but reduced or eliminated smaller peaks and flattened many valleys, as expected by the more limited range of the free radical diffusion. The no-primary-control TSA labeling produced a nearly flat map. In a separate experiment, incubating cells with free HRP to produce a uniform concentration of tyramide free radicals also produced a nearly flat map (not depicted). Together, these two controls indicate the absence of significant variation in the accessibility of DNA in different types of chromatin to reaction with the tyramide free radicals.

Valleys in the SON TSA-Seq map align with lamina-associated domains (LADs) identified by lamin B1 DamID (Fig. 2 B), consistent with the known depletion of nuclear speckles near the nuclear periphery (Carter et al., 1991). Unlike the relatively abrupt transitions between inter-LADs and LADs defined by the DamID molecular proximity mapping method, SON TSA-Seq maps typically show gradual, approximately linearly decreasing signals over the several Mbps separating SON TSA-Seq peaks and valleys, which then levels close to the corresponding lamin B1 inter-LAD to LAD boundary (Fig. 2 B, red boxes). We obtained nearly identical TSA-Seq maps using a different anti-SON antibody (Ab2; see Materials and methods) as well as a monoclonal antibody to a different speckle marker (phosphorylated SC35; Fig. S1, B and C). These speckle TSA-Seq maps are distinctly different from published SC35 and SON ChIP-Seq results, showing near uniform levels at Mbp-scale resolution and localized peaks over cis-regulatory regions at the resolution level of individual genes (Fig. S1 D; Ji et al., 2013; Kim et al., 2016). In contrast with TSA-Seq, which measures cytological distance to the staining target regardless of whether the proteins are within cross-linking distance, ChIP-Seq as a molecular proximity method detects only the small fraction of SC35 or SON likely associated over nascent transcripts within cross-linking distance to DNA. Indeed, these published ChIP-Seq datasets show localized peaks over all active genes regardless of their distance to nuclear speckles (Fig. S1 E).

SON TSA-Seq score is proportional to cytological distance from nuclear speckles and can be calibrated to estimate mean distance in micrometers

We used 3D immuno-FISH to validate that SON TSA-Seq scores are proportional to distance from nuclear speckles. We measured the distances to the nearest speckle using eight BAC FISH probes corresponding with three peaks, three valleys, and two transition regions as visualized in SON TSA-Seq color-coded decile and genomic plots (Fig. 2, C and D; and Fig. S2, A–C, probes A–H). A strikingly deterministic localization near nuclear speckles was seen for probes A–C, corresponding with large SON TSA-Seq peaks. Nearly 100% of alleles probed mapped within 0.5 µm of a speckle with mean distances between ∼0.1–0.2 µm. Probes from transition (probes D and E) and valley (probes F–H) regions showed progressively larger mean speckle distances and broader distance distributions with decreasing SON TSA-Seq scores (Fig. 2, D and E; and Fig. S2, A–C, probes A–H).

If the SON TSA signal is dominated by spreading of tyramide free radicals from nearby speckles, then the TSA-Seq dependence

of biotinylated DNA enrichment as a function of speckle distance should follow a similar exponential fit as observed by microscopy of TSA staining (Fig. 1 B). Indeed, we can closely fit an exponential curve (Fig. 3 A, red lines) to the plot of DNA genomic enrichment versus mean speckle distance for DNA probes A–H (black squares). For each TSA staining condition, the estimated decay constants (R_0) from genomic enrichment plots matched the decay constant calculated from the microscopy of TSA-labeled lac operator sites within the estimated fitting error (Fig. S2 D). To test how well this exponential fit would apply to other genomic regions, we acquired new FISH speckle distance measurements for an additional 10 BAC FISH probes targeting different genomic regions (Fig. S2, A–C, probes I–R). Plotting mean speckle distance versus TSA genomic enrichment for these additional FISH probe regions showed a close match to the original fitted curve (Fig. 3 B, open circles; 0.039 [0.035] µm mean [median] difference between measured versus predicted distances).

To further test how well TSA-Seq could estimate mean distance genome-wide, we used these exponential fits to convert TSA enrichment to an estimated mean distance (Fig. 3 C). We observe a very high reproducibility across the entire genome comparing calculated distances from condition 1 versus condition 2 SON TSA-Seq (residual mean, median, SD: 0.060, 0.043, and 0.063 µm, respectively; Fig. 3 D). Reproducibility is also high for condition 3 for predicted mean speckle distances less than ∼0.4 µm (Fig. 3 C). Above 0.4 µm, speckle distance estimation is noisy because the SON TSA genomic enrichment decay curve has reached a plateau equal to background (Fig. 3 A), which also explains the flattening of valleys observed in condition 3 SON TSA-Seq genomic plots (Fig. 2 B). Interestingly, the DNA replication timing profile is strikingly correlated with the mean speckle distance map (Fig. 3 C).

Applying TSA-Seq to other nuclear compartment markers provides an integrated view of nuclear organization that mirrors microscopic view

To further test the potential for TSA-Seq to probe nuclear organization, we performed TSA-Seq against additional nuclear compartment markers. We chose lamin A and lamin B (LB1 and LB2) as targets to further compare the relationship between chromosome distances to the nuclear lamina versus speckles.

Lamin A and B TSA-Seq signals are very similar and strikingly inversely correlated with speckle TSA-Seq signals throughout most of the chromosome regions, i.e., peaks, valleys, and the linear trajectories connecting peaks with valleys (Figs. 4 and 5 A). Whereas the lamin B1 DamID signal shows a more binary distribution, lamin TSA-Seq maps show distinctly different valley depths for different inter-LAD regions. Deeper valleys are correlated with larger SON TSA-Seq peaks (Figs. 4 B and 5 A, boxes). The genome-wide inverse correlation between lamin and speckle TSA-Seq percentiles defining a nuclear lamina–speckle axis is demonstrated by the 2D scatterplot of lamin B versus SON TSA-Seq (Fig. 5 B).

It is important to note that our quantitative model for how TSA-Seq works predicts that any speckle marker with similar immunofluorescence speckle staining as SON, regardless of whether they are colocalized at a molecular level, should show
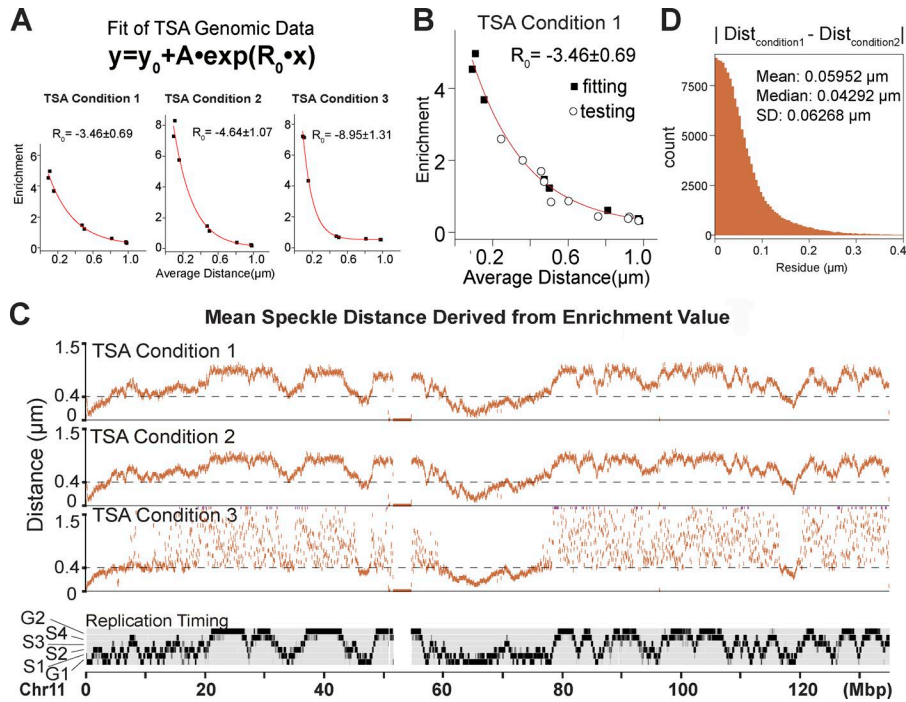
Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology   4029
https://doi.org/10.1083/jcb.201807108

Figure 3. **Deriving an accurate genome-wide mean distance map to nuclear speckles from the SON TSA-Seq score. (A)** Exponential fit (red line) of SON TSA-Seq fold-enrichment (conditions 1–3) versus average speckle distances measured by FISH for eight probes (black squares). **(B)** Enrichment and distances measured for 10 additional FISH probes (open circles) map close to the fitted curve calculated using the original eight probes (shown for condition 1). **(C)** Top to bottom: Speckle distances (Chr11) calculated from the inverse of exponential fit (conditions 1–3) versus Repli-Seq DNA replication timing. **(D)** Histogram of genome-wide distance residuals between condition 1– and 2–calculated speckle distances.

a very similar TSA-Seq pattern. Indeed, the signals for the SC35 and SON TSA-Seq are nearly identical (Fig. 5, A and B; and Fig. S1 C). This is also true for lamin A and B (Fig. 5 B).

We also chose RNA pol II–Ser5p, using the monoclonal 4H8 antibody that recognizes this poised form of RNA pol II (Xie et al., 2006), as a general marker for active nuclear domains to compare with the speckle TSA-Seq chromosome patterns. RNA pol II–Ser5p TSA-Seq shows a lower dynamic range and more binary signal compared with SON TSA-Seq (Fig. 5 A). In contrast with SON TSA-Seq triangular-shaped peaks of varying heights, RNA pol II–Ser5p TSA-Seq shows broader, plateau-shaped peaks of similar heights. The correlation between SON and RNA pol II–Ser5p TSA-Seq is considerably lower than the correlations seen between related datasets such as lamin A versus lamin B TSA-Seq or SON versus SC35 TSA-Seq (Fig. 5 B). Instead, the RNA pol II–Ser5p TSA-Seq chromosomal profile strikingly matches the Hi-C eigenvector (Fig. 5 A), delineating the A (positive) and B (negative) Hi-C compartments known to be correlated with inter-LADs and LADs, respectively.

Comparing the TSA-Seq browser views with direct microscopy visualization of DNA, RNA pol II–Ser5p, SON, and H3K27me3 (Fig. 5 C) provides an integrated view of nuclear organization. RNA pol II–Ser5p and nascent transcript foci are largely excluded from a thin zone at the nuclear periphery, but they are distributed throughout most of the nuclear interior, although they surround and may even be concentrated near nuclear speckles (Fig. 5, C and D). The close spacing of these foci and spreading of the TSA staining produces a merged, low-contrast RNA pol II–Ser5p TSA image. This relatively uniform RNA pol II–Ser5p TSA staining pattern (Fig. 5 E)—higher intensity in the nuclear interior and lower intensity in a peripheral zone close to the nuclear periphery (Fig. 5 E)—explains the low and flat RNA pol II–Ser5p TSA-Seq peaks and valleys overlapping with inter-LADs and LADs, respectively (Fig. 5 A).

At higher resolution, large SON TSA-Seq peaks align only with a subset of RNA pol II–Ser5p TSA-Seq–positive regions, consistent with the RNA pol II–Ser5p immunostaining result: although there is a high concentration of RNA pol II–Ser5p that forms "rings" around the speckle periphery (Fig. 5 C, white arrowheads), there are also many RNA pol II–Ser5p foci widely distributed throughout the nuclear interior that are not speckle associated. Larger SON TSA-Seq speckle peaks align with deeper lamin A and B TSA-Seq valleys, consistent with the more interior localization of nuclear speckles. Most of the smaller SON TSA-Seq local maxima including small peaks and peaks within valleys align with the remaining RNA pol II–Ser5p TSA-Seq–positive regions. These smaller SON TSA-Seq local maxima correlate with small dips or shallow valleys in the lamin TSA-Seq. Thus, these active chromosome regions are predicted to protrude smaller distances into the nuclear interior, consistent with the distribution of RNA pol II–Ser5p foci (Fig. 5 C) and the similar distribution of nascent transcripts (Fig. 5 D) within several hundred nm of the nuclear periphery. Interestingly, the RNA pol II–Ser5p-positive regions that contain large SON TSA peaks overlap extensively with A1 Hi-C subcompartment regions defined in GM12878 cells, which have a similar Hi-C profile to K562 cells (Rao et al., 2014). In contrast, the remaining RNA pol II–Ser5p-positive regions, which contain smaller SON TSA-Seq local maxima, are contained within A2 Hi-C subcompartment regions.

Positive lamin B1 DamID regions (LADs) and lamin TSA-Seq peaks overlap with B2 and B3 repressive Hi-C subcompartments. In contrast, polycomb-associated, repressive B1 Hi-C subcompartments are interspersed between the two types of active Hi-C subcompartments and, to a lesser extent, between B2 and B3. This B1 subcompartment genomic distribution is consistent with the microscopic interspersed foci of RNA pol II–Ser5p and H3K27me3, a marker of polycomb-repressed chromatin, throughout the nuclear interior and the closer approach of H3K27me3 foci to the nuclear periphery (Fig. 5 C).

Chen et al.
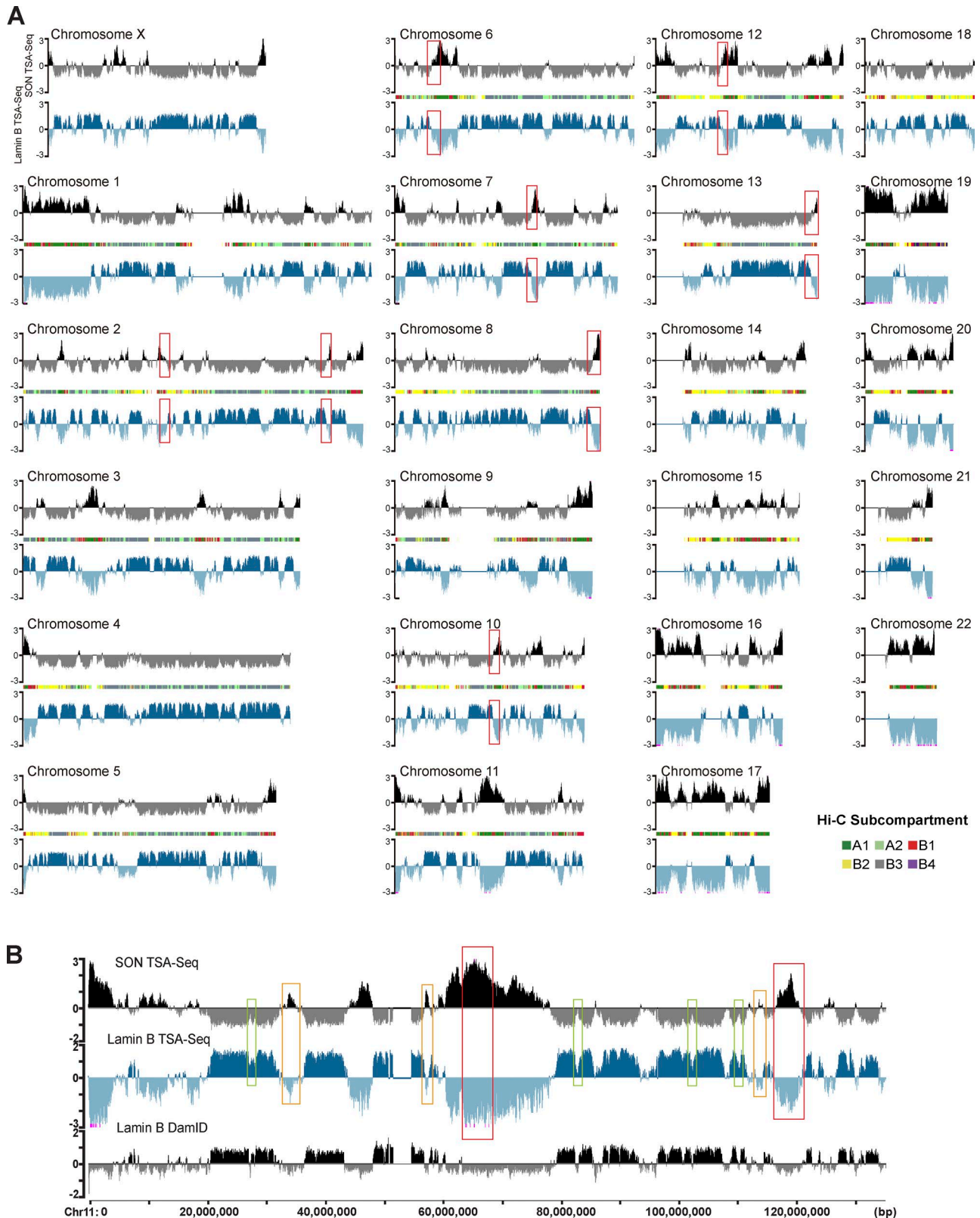TSA-Seq mapping of nuclear genome organization

Figure 4. **Inverse correlation between SON and lamin B TSA-Seq scores as shown for all human chromosomes. (A)** SON TSA-Seq (black) and lamin B TSA-Seq (blue) maps (staining condition 2); y axis shows $\log_2$ ratio of TSA-normalized versus input DNA–normalized reads in 20-kb bins. For comparison, Hi-C subcompartments (GM12878 cells) are shown between the SON and lamin B TSA-Seq plots (color code, bottom right). Plots are based on the hg19 assembly as visualized in the UCSC Genome Browser. Red boxes show examples of linear transitions connecting valleys (peaks) to peaks (valleys) in SON (lamin B) TSA-Seq maps. **(B)** SON TSA-Seq large peaks (red boxes), small peaks (orange boxes), and peaks within valleys (green boxes) correlate with lamin B TSA-Seq large valleys, shallow valleys, and dips between peaks, respectively. Top to bottom: K562 cell SON TSA-Seq, lamin B TSA-Seq, and lamin B1 DamID.
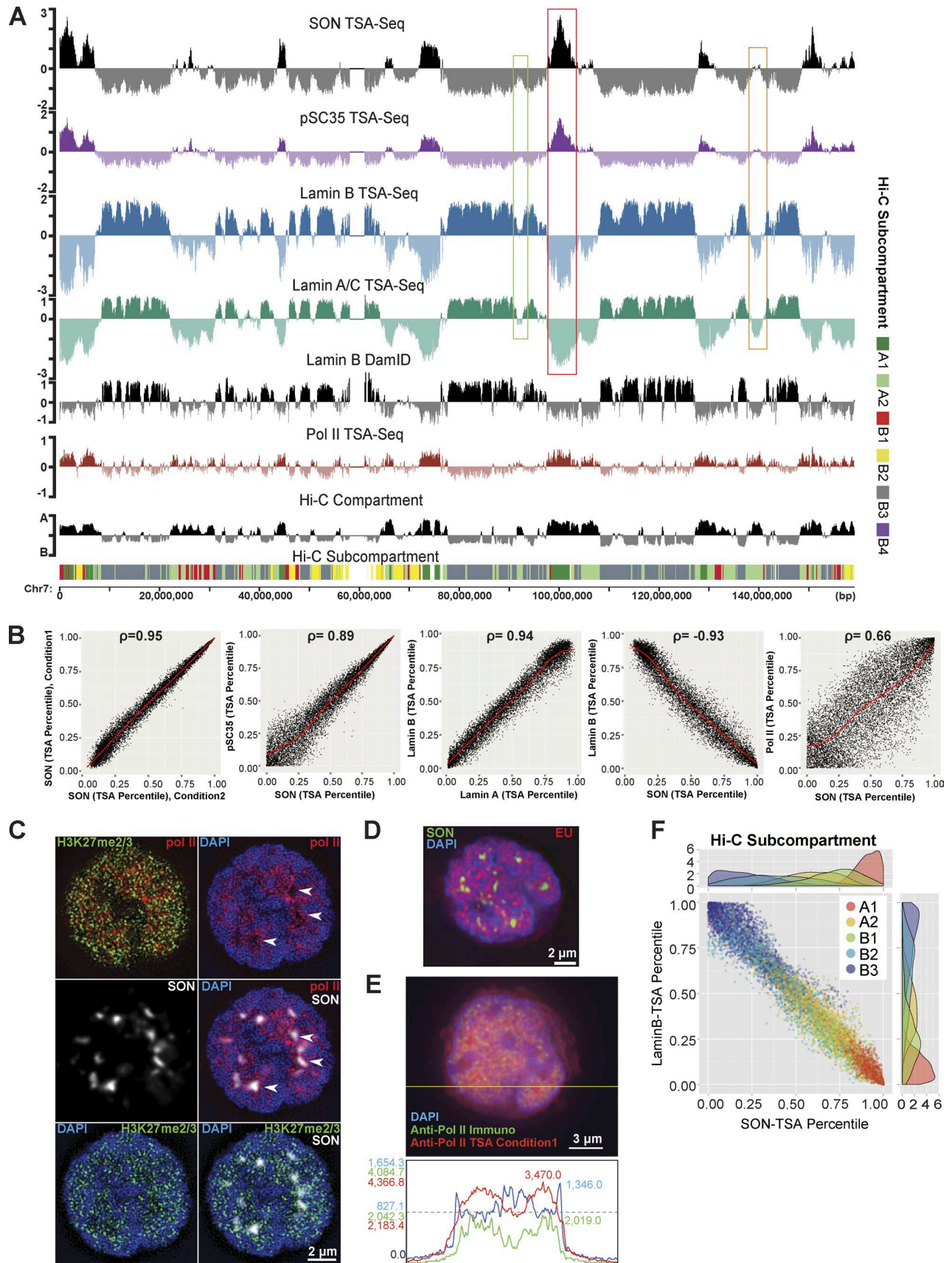
Figure 5. **TSA-Seq versus light microscopy view of nuclear organization. (A)** Top to bottom: TSA-Seq (Chr7) for SON, phosphorylated SC35, lamin B, lamin A/C, lamin B1 DamID, RNA Pol II-Ser5p TSA-Seq, Hi-C compartment eigenvector, and subcompartment assignment. Speckle TSA-Seq large peaks (red boxes), small peaks (orange boxes), and peaks within valleys (green boxes) align with lamin TSA-Seq deep valleys, shallow valleys, and dips between peaks,

A genome-wide view of the relative nuclear spatial distributions of these different Hi-C subcompartments along a nuclear lamina–speckle axis is provided by 2D scatterplots in which each dot, representing a 300-kb chromosome segment, is plotted as a function of lamin B and SON TSA-Seq percentiles (Fig. 5 F). Interestingly, we found a more peripheral distribution of B3 versus B2 subcompartments, hinting at a heterogeneity within or between LADs.

TSA-Seq predicts intranuclear chromosome trajectories, linking genomic analysis with microscopy investigations of nuclear organization

Microscopy-based investigations of intranuclear chromosome topology have previously been limited by their low throughput. In this study, we show the ability of TSA-Seq to predict chromosome trajectories from genome-wide data and thus guide microscopy-based examination of chromosome organization.

The several Mbp genomic distances between SON TSA-Seq valleys and peaks (Fig. 2 B, red boxes) are consistent with the micrometer-scale distances of nuclear speckles from the nuclear periphery and the previously observed ~1–3 Mbp per micrometer compaction of large-scale chromatin fibers (Hu et al., 2009). We predicted that the linear transitions between neighboring peaks and valleys seen in the SON TSA-Seq log2 scores correspond with linear chromosome trajectories that stretch from the nuclear periphery to a nuclear speckle.

To test that the TSA-Seq map can indeed predict both chromosome location and chromosome topology, we first selected a 4-Mbp transition from SON TSA-Seq valley to peak for validation by immuno-FISH. Using two probes located near the LAD boundary (probe 1) and mid-transition (probe 3), we observed 77% of alleles showing probe 1 (red) near the nuclear periphery and probe 2 (yellow) toward the interior, with 96% of examples showing an obvious peripheral to interior orientation for probe 1 versus probe 3 (Fig. S3 A). Using four probes (red) spanning from the LAD boundary to within 1 Mbp of the speckle peak, we observed curvilinear trajectories typically beginning near the periphery and pointing toward a nuclear speckle (Fig. S3 B, green). Adding a fifth probe (yellow) located at the SON TSA-Seq peak showed that this trajectory terminates at the speckle periphery (Fig. 6 A). Similar trajectories were observed for a second 3-Mbp valley-to-peak transition (Fig. 6 B). A probe set covering a complete 11–12-Mbp valley-to-peak-to-valley SON TSA-Seq signal revealed U or V chromosome trajectories from the nuclear periphery to a speckle and back to the nuclear periphery (Fig. 6 C).

In each of these trajectories, the linear slopes seen in both the SON versus lamin TSA-Seq data appear highly inversely correlated, consistent with TSA-Seq reporting distance from the target compartments. These linear slopes contrast with various readouts of transcription that vary significantly along these same trajectories (Figs. 6 C and S3 C), indicating that any small fraction of SON protein associating with nascent transcripts must be too small to contribute significantly to the TSA-Seq signal, consistent with our previous demonstration of using the TSA signal exponential decay to accurately predict the mean speckle distance of chromosome regions (Fig. 3).

Exceptionally deterministic positioning of specific chromosome regions at nuclear speckles

The radial positioning of different chromosome regions is thought to be stochastic (Takizawa et al., 2008; Bickmore, 2013). In contrast, using DNA FISH, we observed that ~100% of the alleles from three different chromosome regions being probed, all belonging to the SON TSA-Seq top 5% percentile, map within 500 nm to nuclear speckles (Fig. 2, D and E). This level of deterministic positioning of a chromosome region relative to a defined nuclear compartment is exceptional. We thus define chromosome regions with SON TSA-Seq scores in the top 5% as speckle-associated domains (SPADs), which have a calculated mean distance to nuclear speckles of <0.32 µm (Fig. 3). Extrapolating from our FISH results, we predict most other SPADs would also map deterministically near nuclear speckles.

Nuclear lamina and speckles as anchor points for chromosome nuclear positioning

Single-cell FISH analysis reveals that SPADs are deterministically localized near speckles but show large variability in nuclear lamina distance (Fig. 7 A, probe C). Conversely, probes from SON TSA-Seq valleys (lamin TSA-Seq peaks) showed a high fraction of alleles positioned near the nuclear lamina but broad nuclear speckle distance distributions (Fig. 7 B, probes Q and R). Probes neither close to lamina nor close to speckles have broad distributions in both speckle and lamina distances, as observed in single cells by DNA FISH. The variable distances of SPADs to the lamina and LADs to speckles reflects the variable intranuclear spatial distribution of speckles. Thus, the nuclear lamina and speckles act to "anchor" LADs and SPADs, respectively, with variable radial positioning of intermediate and speckle-associated sequences. Instead, the linear nuclear lamina–speckle axis revealed by 2D scatter plots of lamina versus speckle TSA-Seq represents an ensemble average over the cell population.

Variation in genomic and epigenomic features as a function of speckle and lamin TSA-Seq scores

We next examined how genomic and epigenomic features distribute with respect to relative distances from nuclear speckles

respectively. **(B)** 2D TSA-Seq percentile scatter plots for (left to right) same target SON but two TSA labeling conditions, SON versus pSC35, lamin A/C versus lamin B, lamin B versus SON, and RNA Pol II–Ser5p versus SON TSA (300-kb bins). **(C)** Top: 3D SIM of RNA Pol II–Ser5p (red) versus H3K27me2/3 (green; left) and RNA Pol II–Ser5p (red) relative to DAPI (blue; right). Middle: SON widefield image (left) superimposed on 3D SIM RNA Pol II–Ser5p (red) and DAPI (blue) images (right). Bottom: 3D SIM DAPI (blue) versus H3K27me3 (green; left) and both DAPI and H3K27me3 superimposed on a widefield SON image (white; right). Arrowheads point to nuclear speckles surrounded by RNA Pol II–Ser5p foci. **(D)** 5-Ethynyl uridine 5-min pulse label showing nascent RNA (red) versus DNA (DAPI) staining and SON immunofluorescence (green). **(E)** Image (top) and aligned intensities (bottom) along line scan (yellow; top) of RNA Pol II–Ser5p immunostaining (green), TSA staining (red), and DAPI (blue). RNA Pol II–Ser5p TSA signal (red) is high over nuclear interior but lower near nuclear periphery (and nucleoli). **(F)** 2D lamin B–SON TSA-Seq scatter plots showing distribution of Hi-C subcompartments (320-kb bin size).
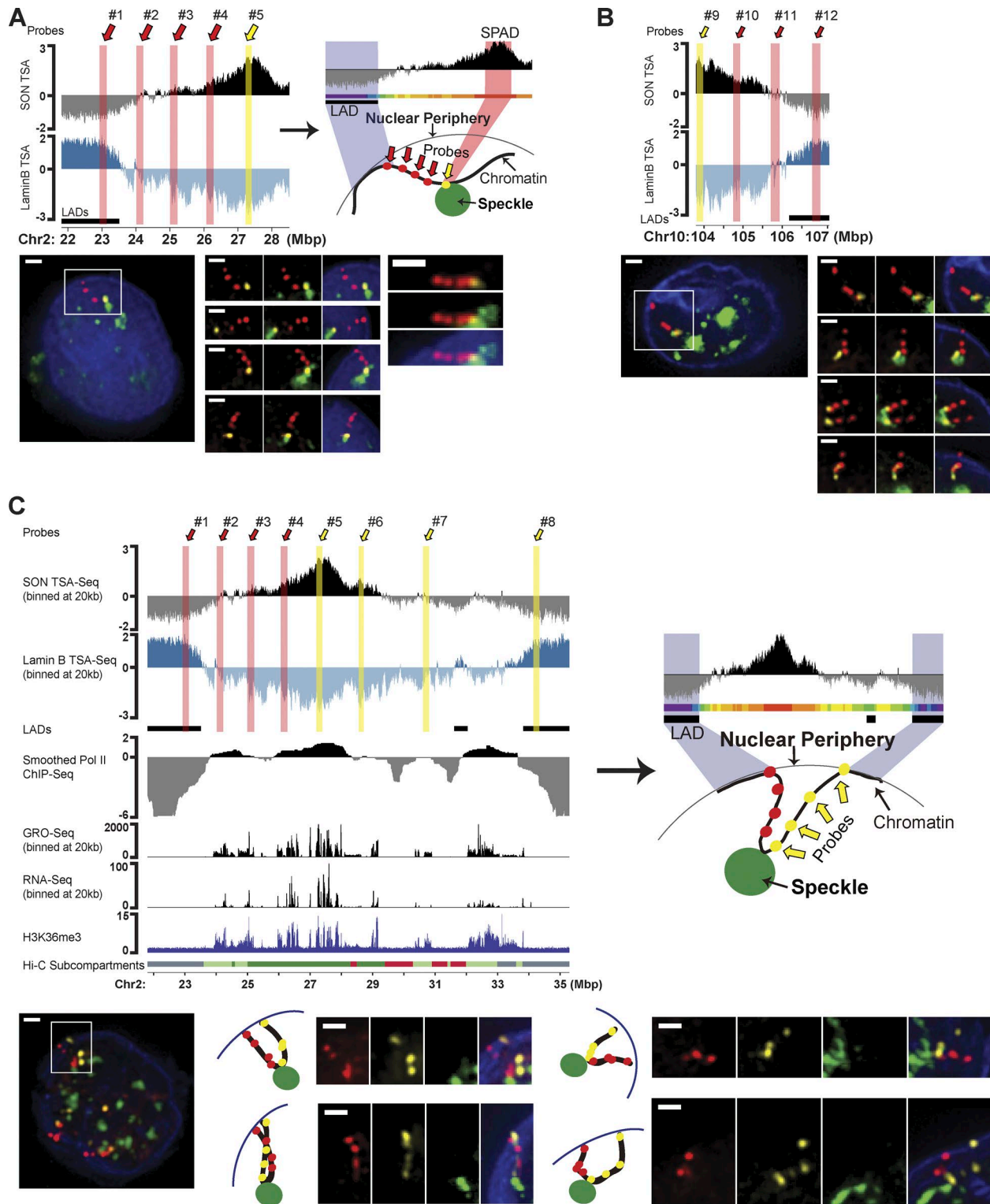
Figure 6. **TSA-Seq predicts chromosome trajectories between the nuclear periphery and speckle. (A)** Immuno-FISH against a 4-Mbp trajectory connecting SON TSA-Seq valley and peak. Top left: SON and lamin B TSA-Seq genomic plot and FISH probe locations. Top right: Schematic of TSA-Seq signal versus chromosome trajectory. Bottom: Probes 1–4 (red) and probe 5 (yellow) FISH and DAPI (blue). Bottom left: Boxed region over FISH nuclear signal. Bottom middle: Trajectory viewed in different channels and merged image (top); three additional examples (bottom). Bottom right: Fifth example. **(B)** Immuno-FISH against 3 Mbp trajectory connecting SON TSA-Seq peak and valley. Top: TSA-Seq genomic plots and FISH probe locations. Bottom left: Probe 9 (yellow) and probes 10–12 (red) FISH and lamin B (blue). Boxed region over FISH nuclear signal. Bottom right: Boxed region viewed in different channels (top) plus three additional examples (bottom). **(C)** Immuno-FISH against ~11 Mbp trajectory from SON TSA-Seq valley to peak to valley. Top left (top to bottom): Probe locations and SON TSA-Seq, lamin B TSA-Seq, smoothed, normalized RNA Pol II ChIP-Seq, RNA-Seq, GRO-Seq (20-kb bins), H3K36me3 ChIP, and Hi-C subcompartments. Right: Schematic of TSA-Seq signal versus chromosome trajectory. Bottom: Probes 1–4 (red) and probes 5–8 (yellow) FISH over lamin B (blue) staining. Bottom left: Boxed region over FISH nuclear signal. Bottom middle and right: different channels (top and middle) plus three additional examples. Bars, 1 µm.
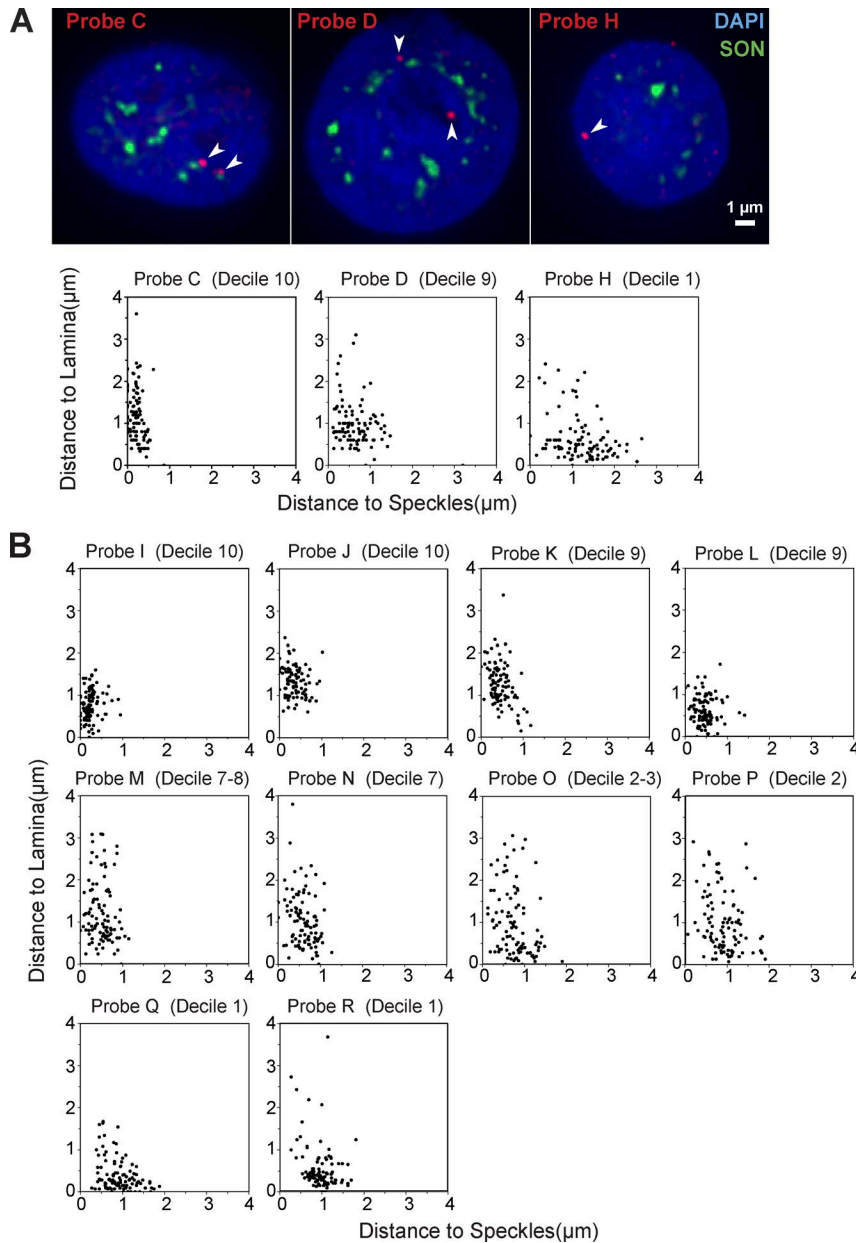
Figure 7. **Single-cell analysis reveals that genomic regions positioned deterministically near nuclear speckles or lamina show stochastic positioning relative to the other compartment. (A and B)** Although 2D lamin-SON TSA scatter plots show a tight inverse correlation, 3D immuno-FISH reveals that in single cells, cytological distances of a given locus from the nuclear speckle or lamina are not tightly correlated. **(A)** Top: DNA FISH probe location (red; arrowheads) relative to nuclear speckles (green) and nuclear DNA staining (blue, DAPI). Probes C (left; decile 10), D (middle; decile 9), and H (right; decile 1; probes same as in Figs. 2, 6, and S2) are from a SON TSA-Seq peak, transition zone, and valley (overlapping with a LAD), respectively; deciles refer to SON TSA-Seq. Bottom: 2D scatter plots showing shortest distances in micrometers from both the nearest nuclear speckle (x axis) and nuclear periphery (y axis) for probes C, D, and H measured for 100 alleles. Probe C shows a very tight distribution of distances (∼0–0.5 µm) close to the nuclear speckle but a broad distribution of distances to the periphery. Instead, Probe H shows a relatively tight distribution of distances for most alleles from the nuclear periphery (∼0–0.8 µm) but a broad distribution of distances to nuclear speckles. Probe D shows distance distributions with an intermediate degree of scatter relative to both nuclear speckle and periphery. **(B)** 10 additional 2D scatter plots for probes I to R (probes same as in Figs. 2, 6, and S2).

and lamin as estimated by TSA-Seq. We first calculated histograms showing segregation of different features as a function of SON TSA-Seq deciles with the 1st and 10th deciles corresponding with the 10% lowest and highest scores, respectively (Fig. 8, A–C; and Fig. S4).

We found a roughly binary division for heterochromatin-related marks: LADs and H3K9me3 are enriched in low SON TSA-Seq deciles and depleted in high SON TSA-Seq deciles; H3K27me3, marking facultative heterochromatin, showed the inverse binary relationship (Fig. 8 A). H4K20me1, alternatively associated with silent (Sims et al., 2006) versus active chromatin (Wang et al., 2008), also shows a binary pattern similar to H3K27me3 (Fig. S4 C). In contrast, nearly all features associated with gene expression showed a linear or steeper gradient spanning all SON TSA-Seq deciles (Fig. 8, B and C; and Fig. S4 C). This includes enhancers, with super-enhancers showing a noticeably steeper gradient than normal enhancers (Fig. 8 D), indicating enrichment

of super-enhancers at the speckle periphery. Gene expression and gene density, accompanied by decreased gene length, also show progressive increases with SON TSA-Seq decile (Figs. 8 C and S4, A and B). In Table S3, we list all protein-coding genes with their corresponding SON TSA-Seq decile and expression levels in K562 cells.

2D scatterplots show that genes with higher expression levels typically show both higher SON and lower lamin TSA-Seq percentiles (Fig. 8, E and F; and Fig. S5 A), while genes with intermediate expression levels show broader distributions along this lamin–SON TSA-Seq axis (Figs. 8 E and S5 A). The most highly expressed genes show a pronounced shift away from the nuclear lamina and toward nuclear speckles (Fig. 8, C and F; and Fig. S5 A), with >50% of the top 5% expressing genes mapping to the top SON TSA-Seq decile (Fig. 8 C, right); moreover, their distribution is disproportionally skewed toward being closer to nuclear speckles as compared with furthest from the nuclear lamina,
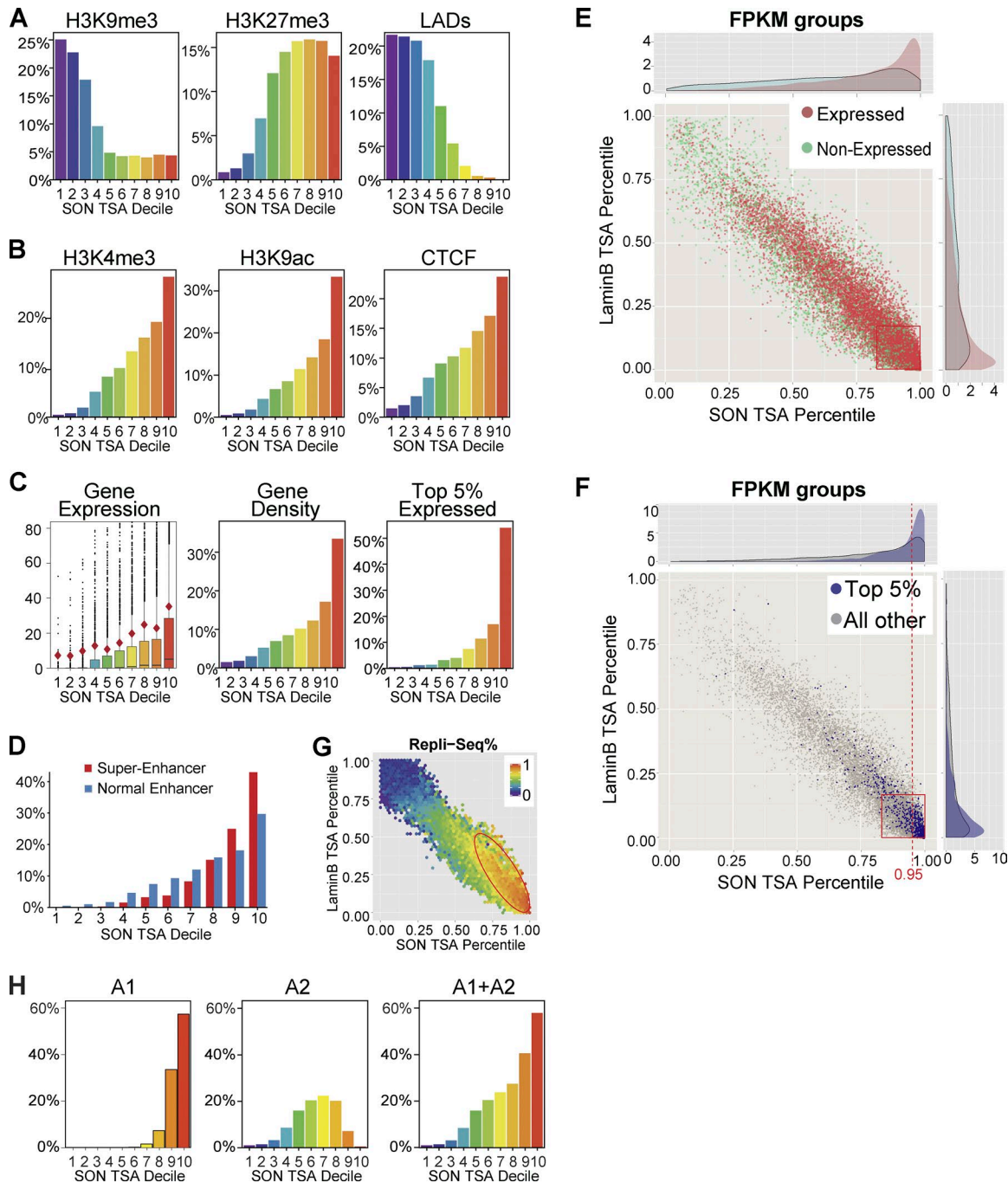
Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology   4035
https://doi.org/10.1083/jcb.201807108

Figure 8. **Gradients in transcription-related marks, gene expression, and DNA replication timing along the speckle–lamin TSA-Seq axis reflect ensemble averaging over cells and chromosome regions. (A–D)** Gene expression and chromatin marks as functions of SON TSA-seq deciles. **(A)** H3K9me3, H3K27me3, and LAD fractions. **(B)** H3K4me3 (peaks), H3K9ac, and CTCF (peaks). **(C)** Protein-coding gene expression (FPKM; left), gene density (middle), and percentage of top 5% active genes (right). Box plot (left) shows median (black line), mean (red diamond), 75% (box top), whiskers equal to 1.5× box size, and outliers (black dots). **(D)** Super-enhancers versus enhancers (percentage). **(E and F)** Lamin B versus SON TSA-Seq percentile 2D scatterplots (projections at top and right). **(E)** Expressed (red) versus nonexpressed (green) protein-coding genes. **(F)** Top 5% expressed genes (blue) versus remaining expressed protein-coding genes (gray). Top 5% expressed genes furthest from nuclear lamina (right; red box) show disproportionate skewing toward speckles (>95% SON TSA-Seq percentile; dashed line). **(G)** DNA replication timing (160-kb bin size): earliest to latest (red to blue). Red ellipse indicates slight skewing of early replicating regions toward higher SON TSA versus lower lamin B TSA percentiles. **(H)** Distributions of A1 (left), A2 (middle), and A1 plus A2 (right) active Hi-C subcompartments as a function of SON TSA-Seq deciles.

particularly for the SPADs (Fig. 8 F, dashed line). Thus, a closer mean distance to speckles rather than a larger distance from the lamina better describes the spatial positioning of the most highly expressed genes.

Similarly, DNA replication timing also distributes along this 2D linear speckle–lamin axis: The earliest replicating chromosome regions cluster nearly uniformly toward lower lamin and higher SON TSA-Seq scores (Figs. 8 G and S5 B), with only a slight

Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology
https://doi.org/10.1083/jcb.201807108

4036

skewing of earlier replicating regions toward disproportionally higher SON TSA-Seq percentiles and intermediate replicating regions toward lower lamin B TSA-Seq percentiles (Figs. 8 G and S5 B, red ellipses).

Notably, the observed linear gradient in transcription-related features and gene expression as a function of distance to nuclear speckles (Fig. 8, B–D) represents an ensemble averaging over different and noncontiguous chromosome regions. For example, the observed linear gradient in transcription-related features and gene expression in the 1D SON TSA decile plots can be deconvolved as a superposition of the distributions of A1- and A2-active Hi-C subcompartment regions (Fig. 8 H).

A subset of transcriptionally active "zones" map close to nuclear speckles

Further comparison of SON TSA-Seq with transcription-related features and gene expression in the context of chromosome neighborhoods reveals two types of transcription hot zones (Fig. 9 A). We define transcription hot zones as Mbp-sized regions with high RNA pol II occupancy, transcription-related marks, and gene expression, where pol II occupancy is measured by a smoothed and normalized RNA pol II ChIP-Seq signal (generated using the 8WG16 antibody recognizing total RNA pol II; see Materials and methods), and transcription and gene expression are measured by global run-on sequencing (GRO-Seq), RNA-Seq, and H3K36me3 ChIP-Seq. These transcription hot zones are also high in gene density. Strikingly, each SON TSA-Seq local maximum localizes near the center of each of the transcription hot zones, defined computationally using the local maxima of the smoothed RNA pol II ChIP-Seq signal (Fig. 9 A, red tick marks). We define these SON TSA-Seq local maxima as the transcription hot zone "peaks."

Transcription hot zone peaks show a noticeably larger variation in the magnitude of the SON TSA-Seq peak value than the variation in magnitude seen for peaks of either the RNA pol II ChIP-Seq or RNA pol II–Ser5p TSA-Seq. The centers of transcription hot zones containing the largest SON TSA-Seq peaks appeared to map to A1 Hi-C subcompartments (Fig. 9 A, red arrows), while the centers of transcription hot zones containing smaller SON TSA-Seq local maxima appeared to map to A2 Hi-C subcompartments.

We created a histogram of SON TSA-Seq peak amplitudes using percentile values at the centers of these transcription hot zones (Fig. 9 B). The observed bimodal distribution suggests the existence of two classes of transcription hot zones: Type I contains large SON TSA-Seq peaks centered over A1 Hi-C subcompartment regions, which are typically flanked by a subclass of A2 Hi-C subcompartment regions (Fig. 9 A, red arrows). Type II contains smaller SON TSA-Seq peaks and peaks within valleys centered over A2 Hi-C subcompartment regions, which are flanked on both sides by repressive Hi-C subcompartment regions (Fig. 9 A, yellow arrows).

Topologically, the SON TSA-Seq local maxima correspond with the most interiorly located regions of a local chromosome trajectory as speckle and lamin TSA-Seq scores are inversely correlated (Figs. 4 and 5). Mean genomic distances between the peaks of transcription hot zones and their nearest LAD are larger for type I versus type II transcription hot zones (Fig. 9 C). These distance distributions suggest a degree of genomic hard wiring of these two types of transcription hot zones contained within simple inter-LADs (Fig. 9 D): type I transcription hot zones span A1 subcompartment regions and extend into flanking A2 subcompartment regions, are longer in length, protrude further into the nuclear interior, and have apexes predicted by their large SON TSA-Seq peak scores to localize adjacent to nuclear speckles (mean apex distance to speckle, 0.36 µm). All transcription hot zones in SPADs are type I, but ~60% of type I hot zones lie outside of SPADs as their SON TSA-Seq percentile is <95%. Type II transcription hot zones span single A2 chromatin domains, are shorter in length, protrude smaller distances into the nuclear interior, and have apexes predicted by their smaller SON TSA-Seq peak scores to localize at intermediate distances to nuclear speckles (mean apex distance to speckle, 0.59 µm).

Transcription hot zones near nuclear speckles have higher overall gene density, a significantly higher number of the most highly expressed genes, and are enriched in housekeeping genes and genes with low transcriptional pausing

Besides their different spatial localization relative to the nuclear lamina and speckles, what other differences might distinguish these two types of transcription hot zones? We plotted the average gene density as a function of genomic distance from peaks of transcription hot zones (Fig. 10 A). Total gene densities and the top 10% expressed gene densities are ~2–2.5-fold higher in type I versus type II transcription hot zones. Both types of transcription hot zones have approximately twofold higher densities of the top 10% expressed genes within 100 kb versus >200 kb of their peak locations. Whereas type I transcription hot zones also show total gene density and the density of moderately expressed genes peaking near their centers, type II transcription hot zones have near constant total gene and moderately expressed gene densities over the 500-kb regions flanking their peaks. These results indicate that transcription hot zones near nuclear speckles have higher overall gene density and a significantly higher number of the most highly expressed genes.

We did not discover any Gene Ontology terms special for genes near nuclear speckles (unpublished data). However, we did see a significant enrichment of genes with low RNA pol II pausing index for speckle-associated type I transcription hot zones within 100 kb of their peaks (Fig. 10 B). We also see a relative enrichment of housekeeping versus nonhousekeeping genes close to nuclear speckles (Fig. 10 C).

Discussion

In this study, we present TSA-Seq, the first genome-wide method capable of directly and quantitatively estimating mean cytological distances from particular nuclear subcompartments. TSA-Seq fundamentally differs from molecular proximity methods such as ChIP-Seq and DamID, which report only on the small fraction of the target protein in close proximity to DNA but are relatively insensitive to much larger concentrations of the same target protein that are not in direct molecular contact distance of the DNA. TSA-Seq instead converts an exponential decay of tyramide free radicals
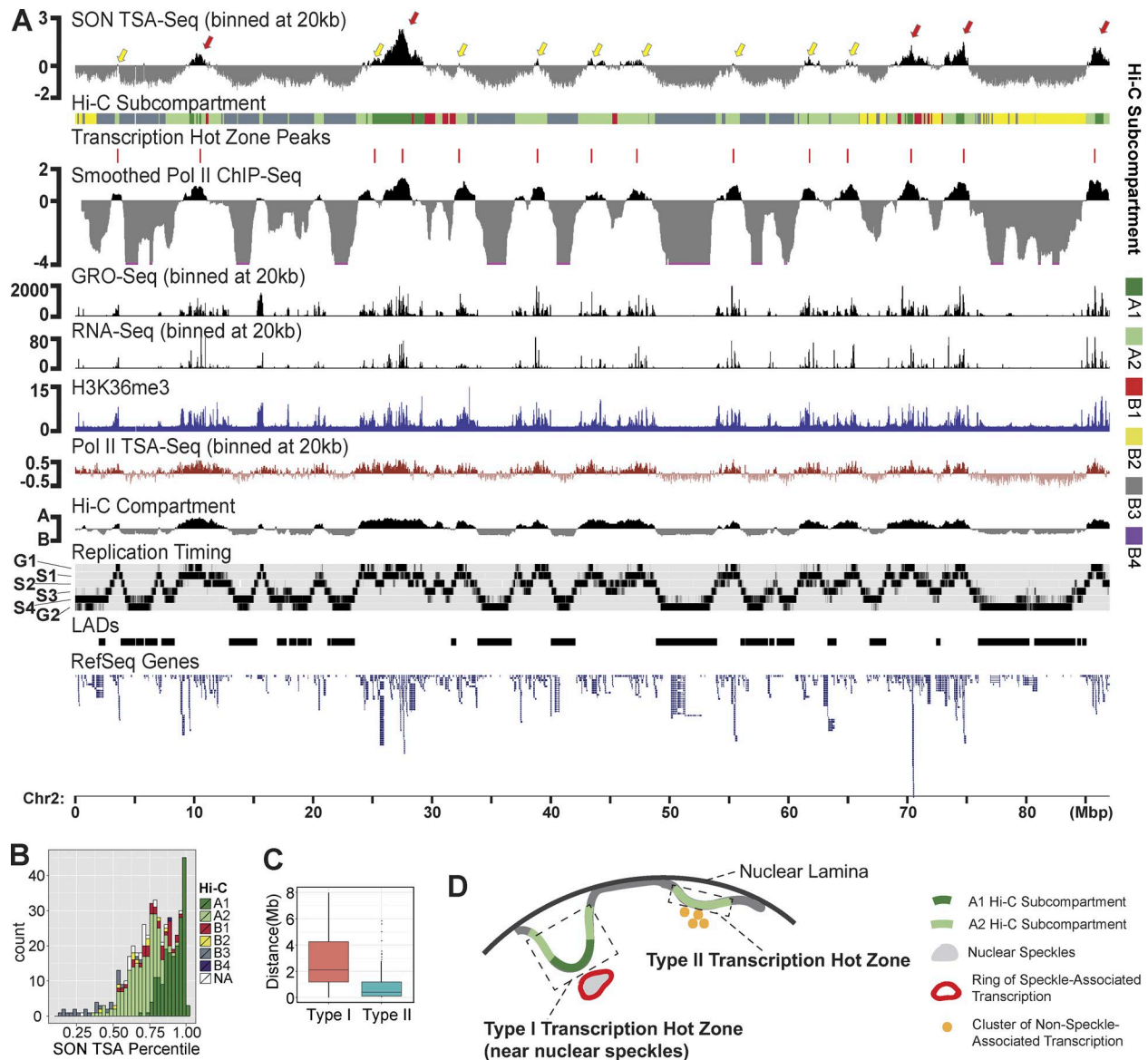
Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology   4037
https://doi.org/10.1083/jcb.201807108

Figure 9. **Two types of transcription hot zones align with SON TSA-Seq local maxima. (A)** Peaks of approximate Mbp transcription hot zones marked by smoothed RNA Pol II ChIP-Seq maxima align with SON TSA-Seq maxima (red tick marks). Top to bottom (Chr2): SON TSA-Seq, Hi-C subcompartments, smoothed, normalized RNA Pol II ChIP-Seq, GRO-Seq (20-kb bins), RNA-Seq, H3K36me3, RNA Pol II–Ser5p TSA-Seq, Hi-C eigenvector (compartment), Repli-Seq, LADs, and RefSeq gene annotations. Arrows showing SON TSA-Seq local maxima align with type I (red) versus type II (yellow) transcription hot zones. **(B)** Bimodal distribution of SON TSA-Seq local maxima number (y axis) color coded by A1 (dark green) versus A2 (light green) Hi-C subcompartment identity versus SON TSA-Seq percentile (x axis). **(C)** Box plots of genomic distances between SON TSA-Seq local maxima and nearest LAD/inter-LAD boundary for A1 (red) versus A2 (blue) SON TSA-Seq peaks. **(D)** Two types of transcription hot zones: Type I, whose apexes map adjacent to nuclear speckles, and type II, whose apexes map to intermediate distances from nuclear speckles.

into a cytological "ruler" that, depending on the staining condition, can be used to probe relative distances over an ~0.5–1.5-μm radius from discretely labeled nuclear compartments separated by distances on this scale. In the case of nuclear speckle TSA staining, we demonstrated the capability of converting TSA-Seq scores into actual mean distances with an estimated accuracy of <100 nm.

Using DNA labeling as our TSA readout was likely key to our success in accurately measuring cytological distance in our implementation of TSA-Seq because the reactivity of DNA to tyramide free radicals was near uniform. Previous applications of TSA to genomic analysis used chromatin rather than DNA pull-

down (Carter et al., 2002; Ching et al., 2013). Using RNA TRAP, peaks of high TSA reactivity near the β-globin locus control region measured ~10 kb in width (Carter et al., 2002). This compares with the several-hundred-kb minimal peak width that we observed in this study using SON TSA-Seq. It is difficult to explain this narrow peak of RNA TRAP labeling given the >500-nm radius of TSA labeling described in this same RNA TRAP study versus the high compaction of chromatin generally observed within mammalian somatic nuclei. Instead, we suspect these narrow peaks might reflect variations in TSA reactivity based on the special nonhistone chromatin composition over these
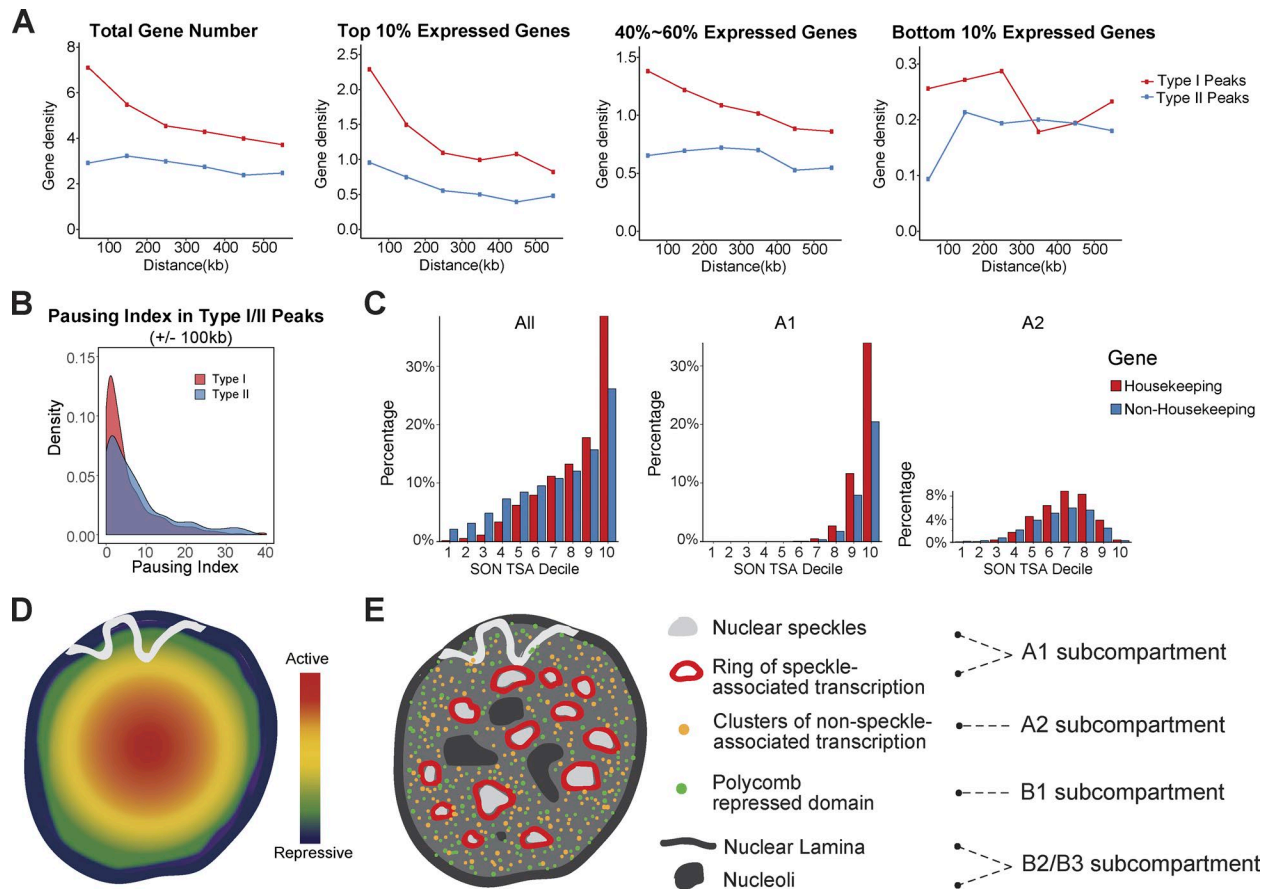
Figure 10. **Differences between type I and II transcription hot zones and a refined model for nuclear organization. (A)** Left to right: Density (y axis; gene number) of total genes, top 10% expressed genes, genes with 40–60% expression levels, and genes with 10% lowest expression as function of genomic distance (x axis; kb) from centers of type I (red) versus type II (blue) transcription hot zones. **(B)** Pausing index distribution for expressed genes within ±100 kb of SON TSA-Seq local maxima contained within type I (A1 peaks) versus type II (A2 peaks) hot zones. **(C)** Percentage (y axis) of housekeeping versus nonhouse-keeping genes versus SON TSA-Seq deciles for all genes (left), genes only in A1 subcompartment regions near speckles (middle), or only in A2 subcompartment regions (right). **(D)** Previous model: Transcriptional activity increases radially from low near the nuclear periphery to high toward the nuclear center. **(E)** New, refined model: Transcriptionally active chromosome regions are depleted from heterochromatic compartments close to the nuclear and nucleolar periphery (black). Apexes of chromosome trajectories protruding into the nuclear interior correspond with centers of transcription hot zones, which localize either at the periphery (red rings) of nuclear speckles (gray; type I) or at interior sites at intermediate distances from speckles, possibly near unknown compartments such as transcription factories (type II). A second repressive compartment enriched in H3K27me3 (green dots) maps throughout the nuclear interior, dispersed between RNA Pol II–Ser5p foci and speckles. Large-scale chromatin fibers (gray curves) traverse between these compartments.

regulatory regions, which were not measured in these previous experiments.

Previous low-throughput FISH research has led to a model in which a gradient of increasing gene expression potential extends from the nuclear periphery to nuclear center (Fig. 10 D; Bickmore, 2013). However, intranuclear spatial positioning of specific genes appeared to be highly stochastic. We propose that previous descriptions of stochastic positioning of chromosome regions as a function of gene expression relative to a nuclear periphery to center axis (Fig. 10 D) conceal a more deterministic positioning of different genomic regions relative to specific nuclear compartments (Fig. 10 E). One such compartment is the nuclear periphery (van Steensel and Belmont, 2017). In this study, we show that nuclear speckles act as a second anchor to organize interphase chromosomal organization.

Indeed, other previous cytological research has suggested that nuclear speckles act as a hub for a subset of active genes and chromosome regions. This included early observations of enrich-

ment of highly acetylated histone H3, the TAF250 HAT, and sites of bromouridine incorporation at the periphery of nuclear speckles (Hendzel et al., 1998). More specifically, this also included the mapping in other cell lines of roughly half of the ~25 active genes examined to the periphery of nuclear speckles (Hall et al., 2006) as well as the mapping of several gene-rich chromosome bands near speckles (Shopland et al., 2003). Our SON TSA-Seq results in K562 cells show these same gene-rich chromosome bands (17q22-24, 17q21, 6p21, and 19q13.3) in regions with high SON TSA-Seq scores.

However, our genome-wide analysis and distance mapping in K562 cells made possible by our TSA-Seq data now has allowed us to build substantially on these earlier findings to propose a refined model of nuclear organization (Fig. 10 E): inter-LADs vary in how far they protrude into the nuclear interior, with centers of transcriptional hot zones enriched in the most highly expressed genes mapping to the most interior location of an inter-LAD chromosome trajectory. These transcription hot zones were divided into two types. Type I transcription hot zones,

Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology 4039
https://doi.org/10.1083/jcb.201807108

mapping very close to nuclear speckles, correspond with the A1 Hi-C subcompartment, have approximately twofold higher gene densities, and are enriched in super-enhancers, the most highly expressed genes, and genes with low transcriptional pausing. Interestingly, factors involved in RNA pol II pause release (pTEFb and cdk12) are enriched in nuclear speckles (Herrmann and Mancini, 2001; Ko et al., 2001; Dow et al., 2010). In contrast, type II transcription hot zones, mapping to A2 Hi-C subcompartments, still correspond with the apexes, or most interior locations, of the inter-LAD loops protruding into the nuclear interior, but they are located at intermediate distances from nuclear speckles.

SPADs, defined as the fraction (~5%) of the genome that associate close to nuclear speckles with near 100% probability, were estimated as having a mean distance to speckle periphery of <0.32 µm. Thus, these SPADs may serve to anchor the most active genomic regions to nuclear speckles, similar to how LADs appear to anchor inactive genomic regions to the nuclear lamina.

Both gene expression and timing of DNA replication align along a nuclear speckle–lamina axis, which we show represents an ensemble average over different cells as well as different types of noncontiguous chromosome trajectories. Our model describes several spatially distinct nuclear compartments: (a) a thin rim adjacent to the nuclear periphery and relatively depleted of RNA pol II and transcription and enriched in LAD regions, likely also localizing at the periphery of nucleoli and centromeres (van Steensel and Belmont, 2017) that were not directly probed in this study; (b) interior regions enriched in A2 Hi-C active subcompartments; (c) intermingled regions enriched in polycomb-silenced regions corresponding with the B1 Hi-C active subcompartment; and (d) the periphery of nuclear speckles where A1, type I transcription hot zones terminate (Fig. 10 E). Our TSA-Seq results in fact assign the differential localization of the Hi-C active A1 and A2 subcompartments to speckle-associated versus interior nuclear locations.

The logic of this nuclear organization remains to be determined. Our model would suggest, however, that chromosome movements of just several hundred nm during cell differentiation or during a cell physiological response might have significant functional significance. For example, a movement of just several hundred nm from the nuclear periphery toward the nuclear interior to sites enriched in RNA pol II foci might be needed to significantly boost the transcription of a gene contained within a repressive chromatin domain. Similarly, a movement of just a few hundred nm from the nuclear interior to adjacent to a nuclear speckle might produce either a significant transcriptional enhancement and/or an enhancement of post-transcriptional processing and/or nuclear export leading to increased gene expression. Indeed, while nuclear speckles contain transcription-related factors, they also contain a larger number of factors involved in different stages of RNA processing and factors related to nuclear export. Thus, any possible changes in gene expression linked to nuclear speckle proximity might result from changes in either transcription and/or post-transcriptional events.

Further analysis to more fully exploit TSA-Seq distance mapping should provide further insights. So far, our analysis has focused on results from a single cell line: K562 cells. One significant technical limitation of TSA-Seq as currently implemented using DNA labeling as the readout is the requirement for hundreds of millions of cells. This is primarily due to the low efficiency of TSA labeling of DNA. However, based on preliminary data, we anticipate future improvements in the TSA-Seq protocol to reduce substantially the number of cells required for the procedure, thereby greatly facilitating future TSA-Seq data acquisition. Additional insights should then come from extension of TSA-Seq to additional cell types and additional nuclear compartments as well as examination of changes in nuclear organization accompanying developmental and cell physiological transitions.

Materials and methods
TSA-Seq
TSA cell labeling
K562 cells were obtained from the ATCC and cultured according to ENCODE Consortium recommendations (http://genome.ucsc.edu/ENCODE/protocols/cell/human/K562_protocol.pdf). SON TSA labeling was done in independent cell batches. For each batch, cells were divided equally and stained according to four labeling protocols: condition 1, condition 2, condition 3, and No 1° control. Identical staining conditions were used except that 50% sucrose was included in the TSA staining buffer for the condition 2 protocol, 0.75 mM DTT in addition to 50% sucrose was added in the TSA staining buffer for the condition 3 protocol, and the anti-SON primary antibody was omitted for the No 1° control protocol. Conditions, cell numbers, and yield for all SON, pSC35, lamin A/C, and lamin B TSA-Seq experiments are summarized in Table S1.

Cells were fixed by adding 10 ml of 8% paraformaldehyde (Sigma-Aldrich) freshly prepared in calcium and magnesium–free (CMF)-PBS to 40 ml cells in culture media, incubating for 20 min at RT, and quenching aldehydes by addition of 5.6 ml of 1.25 M glycine (Thermo Fisher Scientific) and incubation for 5 min at RT. All incubation steps used a rotating platform to mix the cell suspension. Cells were then pelleted by centrifugation and resuspended in CMF-PBS buffer with 0.5% Triton X-100 (Sigma-Aldrich) for 30 min at RT. All centrifugations were for 5 min at 100–200 g. Cells were centrifuged at RT and resuspended in 1.5% H_2O_2 (Thermo Fisher Scientific) in CMF-PBS for 1 h at RT to quench endogenous peroxidases, followed by three washes at RT in PBST (0.1% Triton X-100/CMF-PBS) by repeated centrifugation and cell pellet resuspension.

Cells were resuspended in IgG blocking buffer (5% normal goat serum [Sigma-Aldrich] in PBST) at a concentration of 10^7 cells/ml for 1 h, followed by centrifugation and resuspension in primary antibody solution (rabbit anti-SON primary antibodies [Ab1; HPA023535; Atlas Antibodies] diluted 1:1,000 or 1:2,000 in IgG blocking buffer, rabbit polyclonal anti-SON primary antibodies [Ab2] diluted 1:2,000 in IgG blocking buffer, mouse antiphosphorylated SC35 primary antibodies [S4045; Sigma-Aldrich] diluted 1:500 in IgG blocking buffer, mouse monoclonal anti–lamin A/C primary antibody [clone 5G4] diluted 1:1,000 in IgG blocking buffer, or mouse monoclonal anti–lamin B primary antibodies [clone 2D8] diluted 1:1,000 in IgG blocking buffer) at 10^7 cells/ml, or mouse monoclonal anti-RNA polymerase II primary antibodies (clone 4H8; 05-623; EMD Millipore) diluted 1:300 in IgG blocking buffer at 0.5×10^7 cells/ml and incubated

Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology 4040
https://doi.org/10.1083/jcb.201807108

20–24 h at 4°C. For the no-primary control, cells were incubated in IgG blocking buffer only. Cells were then washed 3× with PBST, resuspended in IgG blocking buffer with secondary antibodies at 10^7 cells/ml, and incubated 20–24 h at 4°C. Secondary antibodies (Jackson ImmunoResearch Laboratories, Inc.) used were HRP conjugated and either goat anti-rabbit secondary antibody diluted 1:1,000 or HRP-conjugated goat anti-mouse secondary antibody diluted 1:200 or 1:1,000.

Cells then were washed 3× in PBST and resuspended in reaction buffer A (CMF-PBS for condition 1, CMF-PBS with 50% sucrose for condition 2, or CMF-PBS with 50% sucrose and 1.5 mM DTT for condition 3) at 2×10^7 cells/ml. Equal volumes of buffer B (8 mM tyramide-biotin stock solution diluted 1:5,000 with 0.003% H_2O_2 in CMF-PBS for condition 1 or CMF-PBS with 50% sucrose for conditions 2 and 3) were then added. Cells were incubated for 10 min and then washed 3× in PBST at RT. Tyramide-biotin was prepared using tyramide hydrochloride (Sigma-Aldrich), EZ-link Sulfo-NHS-LC-biotin (Thermo Fisher Scientific), dimethyl formamide (Sigma-Aldrich), and triethylamine (Sigma-Aldrich) as previously described (Hopman et al., 1998).

Monitoring TSA labeling by antibiotin immunostaining
After each TSA labeling batch, one drop of the cell suspension was placed on a glass coverslip, air dried, and fixed with 1.6% paraformaldehyde in CMF-PBS at RT for 15 min. After washing in PBST 3 × 5 min, coverslips were incubated 30 min at RT with streptavidin–Alexa Fluor 594 (Invitrogen) diluted 1:200 in CMF-PBS, washed 3 × 5 min at RT in PBST, and mounted in antifade media (0.3 µg/ml DAPI [Sigma-Aldrich]/10% wt/vol Mowiol 4–88 [EMD Millipore]/1% wt/vol DABCO [Sigma-Aldrich]/25% glycerol/0.1 M Tris, pH 8.5).

DNA isolation from labeled cells
Cell lysis and protease digestion were done at 55°C with constant mixing in 10 mM Tris, 10 mM EDTA, 0.5% SDS, and 500 µg/ml proteinase K (NEB) for 8–12 h (10^7 cells/ml). NaCl was added for a final 0.2 M concentration, and the lysate was incubated 20–24 h at 65°C with constant mixing to reverse formaldehyde cross-linking. DNA was extracted 2× using an equal volume of phenol (Thermo Fisher Scientific)/chloroform/isoamyl alcohol (25:24:1) and 1× with an equal volume of chloroform/isoamyl alcohol (24:1). Tubes were kept at 55°C for 1 h with the cap off to evaporate phenol residues, and then they were treated with 25 µg/ml RNaseA (QIAGEN) at 37°C for 1 h and extracted again. DNA was then precipitated with 1:10 volume of 3 M Na acetate, pH 5.2, and two volumes of 100% EtOH at –20°C for 20–120 min.

DNA fragmentation
DNA was dissolved in double-distilled H_2O and sonicated to 200–800 bp with the majority of the fragments at ~500 bp using a Branson 450 digital sonifier (with Model 102 converter and microtip) or a Bioruptor (UCD-200 or Pico). For the Branson sonifier, each sonication was done twice with 250 µl volume in 1.5 ml Eppendorf tubes on ice for 20 s (1-s chase, 1-s pulse, and total of 40 s). For the Bioruptor, sonication was done with 30 s on and 30 s off, but the number of cycles varied from tube to tube.

Monitoring biotin labeling of DNA by dot blot
DNA was spotted onto nitrocellulose membrane (0.45 µm; Bio-Rad Laboratories) and UV cross-linked to the membrane (0.24 J; UV Stratalinker 1800; Agilent Technologies). The membrane then was blocked in SuperBlock (TBS) blocking buffer (Thermo Fisher Scientific) with 0.05% Tween-20 (Sigma-Aldrich) for 1 h, incubated 12–16 h at 4°C with streptavidin-HRP (Invitrogen) diluted 1:2,500 in the same blocking buffer, washed at RT with TBST (TBS/0.05% Tween-20), treated with SuperSignal West Pico chemiluminescent substrate (Thermo Fisher Scientific), and exposed with CL-XPosure film (Thermo Fisher Scientific). For TSA labeling with a specific antibody, we estimate biotin labeling of ~0.0004–0.03 biotin per 10 kb depending on the target and TSA condition; based on an average DNA fragment size of ~500 bp, this produced ~0.002–0.15% of fragments labeled with biotin, assuming one biotin per fragment.

Affinity purification of biotinylated DNA
The low level of biotin labeling achieved by the TSA labeling conditions used in this study meant that we needed very large cell numbers (approximately hundreds of millions depending on staining condition) to recover sufficient biotinylated DNA for library construction. Therefore, we performed the TSA staining in multiple batches of cells. DNA derived from TSA staining batches that produced satisfactory TSA labeling, as verified by microscopy and dot blot analysis, was pooled for affinity purification. The required ratio of DNA and Dynabeads M-270 streptavidin (Invitrogen) was estimated from the anti-biotin dot blot results and the known binding capacity of the beads. Beads were washed 3× in wash and binding (W&B) buffer (5 mM Tris-Cl, pH 7.5, 0.5 mM EDTA, and 1 M NaCl) using a magnetic stand. 10% input DNA was set aside. The remaining DNA was mixed with an equal volume of 2× W&B buffer. In the initial protocol, DNA samples in 4× beads' original volume were mixed with the beads and nutated at RT for 0.5 h. The beads were immobilized in a magnet, the solution was removed, and an additional 4× original bead volume of DNA in W&B buffer was added. This was repeated until all DNA from each staining condition had been bound to the beads. The last incubation was extended for 8 h at 4°C. Beads then were washed for 4 min at RT 4× with 1× W&B buffer, 4× with TSE 500 (20 mM Tris, pH 8.0, 1% Triton X-100, 0.1% SDS, 2 mM EDTA, and 500 mM NaCl), 4× with 1× W&B buffer, and 4× with TE buffer (10 mM Tris and 10 mM EDTA, pH 8.0). In a later modified protocol, DNA samples were mixed with the desired amount of beads all at once regardless of the volume of DNA solution, and the mixture was incubated with rotation at RT for 1 h and then 4°C for 8 h. Beads then were washed five times. For each wash, beads were first resuspended in 1 ml 1× W&B buffer with Triton X-100 added at 0.05% final concentration, then were incubated at 55°C on heat block for 2 min, and then were nutated at 55°C for 2 min. Beads were transferred to a new tube after each wash.

After the final wash, beads were resuspended in 10 mM Tris, 10 mM EDTA, 0.5% SDS, and 1 mg/ml proteinase K at 600 µl per 100 µl original bead volume. Free biotin was added using 30 µl of a 2-mM stock solution per 100 µl original bead volume, and beads were digested overnight at 55°C with rotation. DNA was

then extracted using phenol/chloroform as described in the DNA isolation from labeled cells section and was precipitated using 0.05 µg/µl glycogen (Fermentas) as a carrier with 1:10 volume of Na acetate and two volumes of 100% EtOH for 12–16 h at −20°C.

Library preparation and high-throughput DNA sequencing

For each of the TSA-Seq experiment staining conditions, input and streptavidin-pulldown DNA were used to make sequencing libraries using the Kapa Hyper Prep kit (Kapa Biosystems). DNA was blunt ended, 3′-end A tailed, and ligated to indexed adaptors with 6-nt barcodes. PCR amplification selectively enriched for those fragments with adaptors on both ends. Amplification was performed for 8–16 cycles with HiFi polymerase (Kapa Biosystems). A Bioanalyzer DNA high-sensitivity chip (Agilent Technologies) was used to determine library fragment sizes. Library fragment concentrations were measured by quantitative PCR on a CFX Connect Real-Time System (Bio-Rad Laboratories) before pooling and sequencing. Libraries were pooled at equimolar concentration and sequenced for 101 cycles on an Illumina HiSeq 2000, HiSeq 2500, or HiSeq 4000 using a HiSeq Sequencing-By-Synthesis kit, version 4. The raw BCL files were converted into demultiplexed compressed fastq files using the bcl2fastq v1.8.2 and later v2.17.1.14 conversion software (Illumina). 27–81 million 100-nt reads were obtained for each sample with average quality scores >30. Library preparation and sequencing were done by the High-Throughput Sequencing and Genotyping Unit of the Roy J. Carver Biotechnology Center at University of Illinois at Urbana-Champaign (UIUC).

Lamin B1 DamID

DamID of lamin B1 was performed in K562 cells as described (Vogel et al., 2007; Meuleman et al., 2013) except that before the DpnI digestion, 1.5 µg genomic DNA was treated for 1 h with 5 U Antarctic phosphatase (M0289S; NEB) in 20 µl 1× Antarctic phosphatase reaction buffer (NEB) followed by ethanol precipitation. This additional step prevents ligation of the adaptor to DNA fragments originating from apoptotic cells. Samples from two independent biological replicates were hybridized to human genome tiling microarrays (Nimblegen HD2) that contained 2.1 million probes designed as previously reported (Meuleman et al., 2013).

3D immuno-FISH

FISH probes were made from BACs (Invitrogen; See Table S2 for BAC list). Probes were labeled with biotin or digoxigenin (Dig) using nick translation (BioNick kit; Invitrogen) according to the manufacturer's instructions, with a minor change for Dig labeling in which the kit's 10× dNTP mix was replaced by 0.125 mM Dig-11-dUTP (Roche), 0.375 mM dTTP (NEB), 0.5 mM dATP (NEB), 0.5 mM dCTP (NEB), 0.5 mM dGTP (NEB), 500 mM Tris-HCl, pH 8.0, 50 mM $MgCl_2$, and 500 µg/ml BSA (NEB). Alternatively, DNA was end labeled using terminal transferase as described elsewhere (Dernburg, 2011) with the following reagents: terminal transferase and 5× buffer (Fermentas), dATP (NEB), and biotin-14-dATP (Thermo Fisher Scientific) for biotin labeling or dTTP (NEB) and Dig-11-dUTP (Roche) for Dig labeling. Concentrations of labeled and unlabeled nucleotides were 54 and 108 µM, respectively, for a 25-µl reaction with 2 µg fragmented DNA.

K562 cells were plated on poly-L-lysine (70,000–150,000 molecular weight; Sigma-Aldrich)–coated Circes #1.5 glass coverslips (Thermo Fisher Scientific) in a 24-well plate at 0.3 ml per well and 10^5–6×10^5 cells/ml in RPMI culture media plus 10% FBS (Sigma-Aldrich). Cells attached after 15 min at 37°C in a humidified incubator with 5% CO_2. Cells were then permeabilized in 0.1% Triton X-100 in CMF-PBS for 1 min at RT, fixed with freshly prepared 3.6% paraformaldehyde (Sigma-Aldrich) in CMF-PBS for 10 min at RT, and quenched with 0.5 M glycine (Thermo Fisher Scientific) in CMF-PBS for 5 min at RT.

Our 3D immuno-FISH was modified from that previously described (Solovei and Cremer, 2010). For anti-SON and anti–lamin B immunostaining, cells were blocked in IgG blocking buffer for 1 h at RT, incubated with rabbit anti-SON antibodies diluted 1:1,000 with or without mouse anti–lamin B antibodies diluted 1:250 in IgG blocking buffer overnight at 4°C, washed with PBST 3 × 5 min at RT, incubated for 1 h at RT with FITC-labeled secondary goat anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc.), which was diluted 1:500 with or without aminomethyl-coumarin-labeled secondary goat anti-mouse IgG (Jackson ImmunoResearch Laboratories, Inc.) that was diluted 1:50 in IgG blocking buffer, and washed 3 × 5 min with PBST at RT. Cells were then post-fixed with 3% paraformaldehyde in CMF-PBS for 10 min at RT, quenched with 0.5 M glycine in PBS for 5 min at RT, and washed with PBST for 5 min at RT.

Coverslips were incubated in 20% glycerol in CMF-PBS for 30–60 min at RT, frozen, and thawed 6× using liquid nitrogen, washed 3 × 5 min with PBST, rinsed and incubated on ice with 0.1 M HCl/0.7% Triton X-100/2× saline-sodium citrate (SSC) for 10 min, and then washed 3 × 5 min with 2× SSC. FISH probes (100 ng for probes made by nick translation and 20 ng for probes made using terminal transferase) in 4 µl hybridization buffer (10% dextran sulfate, 50% formamide, and 2× SSC) were then added to coverslips and sealed onto glass slides with rubber cement. Samples were denatured on a hot plate at 76°C for 3 min and then hybridized 18–72 h in a humid chamber at 37°C.

Coverslips then were washed at 70°C in 0.4× SSC for 2 min and at RT in 4× SSC for 5 min. Alternatively, they were washed at 37°C for 3 × 5 min in 2× SSC and then at 60°C for 3 × 5 min in 0.1× SSC and at RT in 4× SSC for 5 min. The biotin and Dig signals were detected by incubation for 40–120 min at RT in a humidified chamber with streptavidin–Alexa Fluor 594 diluted 1:200 and/or mouse anti-Dig–Alexa Fluor 647 (Jackson ImmunoResearch Laboratories, Inc.) diluted 1:100 in 4× SSC with 4% BSA (Sigma-Aldrich). Coverslips were washed in 4× SSC/0.1%Triton X-100 3 × 5min and mounted in DAPI-containing antifade mounting media.

Microscopy

3D optical-section images were taken with 0.2-µm z steps using an Applied Precision personal DeltaVision microscope system and DeltaVision SoftWoRx software (GE Healthcare) with a 60×, 1.4 NA objective lens and a CoolSNAP HQ2 charge-coupled device camera. Images were deconvolved using an enhanced-ratio iterative constrained algorithm (Agard et al., 1989) within SoftWoRx. For immuno-FISH images, 3D distances were measured from the center of each FISH spot to the edge of its nearest nuclear speckle using the Standard Two-Point Measure tool in SoftWoRx. 3D sur-

Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology   4042
https://doi.org/10.1083/jcb.201807108

face plots were generated using FIJI software (ImageJ; National Institutes of Health; Schindelin et al., 2012) with the Spectrum lookup table. All figure panel images were prepared using FIJI and Photoshop CS6 (Adobe). Image enlargement used bicubic interpolation in Photoshop CS6.

Exponential fitting of microscopy and TSA-Seq data
For microscopy images, the green- and red-channel pixel intensities along a line were exported using the line profiling function in FIJI software and plotted against distance using OriginPro 9.1 (OriginLab). We used OriginPro 9.1 to fit the TSA signal to the equation $y = y_0 + A * exp(R_0 * x)$, where y is the TSA signal intensity, and x the distance to the intensity peak location, using the iterative Levenberg Marquardt algorithm. All fitting results have adjusted R square > 0.87. OriginPro 9.1 was also used to fit the TSA-Seq genomic data using the same equation and settings as used for the microscopy images. In this case, y is the TSA enrichment ratio between pulldown and input, and x is the mean speckle distance measured from 100 alleles by 3D immuno-FISH. Eight FISH probes (probes A–H) were used to do the fitting, while 10 additional probes (probes I–R) were used to test the goodness of the fit. All fitting results have adjusted R square > 0.97.

Read mapping and processing
Raw sequencing reads were first mapped to the human genome (UCSC hg19) using Bowtie 2 (version 2.0.2; Langmead and Salzberg, 2012) with default parameters. Chromosome Y was excluded in the reference genome to reduce mapping bias since the K562 cell line was derived from a female. PCR duplicate reads were removed using rmdup command from SAMtools (version 0.1.19; Li et al., 2009) with default parameters.

We then normalized and processed the TSA-Seq data using a sliding window approach as follows. First, the normalized TSA-Seq enrichment score was calculated as in Eq. 1, where N_{TSA} is the number of mapped reads in a 20-kb window in the pulldown sample, and T_{TSA} is the total number of mapped reads of pulldown sample (N_{input} and T_{input} are defined similarly for input data). This approach can be interpreted as the fold change between the pulldown sample and the input corrected by the total number of mapped reads in each sample. The window slides on the genome with a step size of 100 bp.

$$\text{TSA-Seq enrichment score} = \log_2\left(\frac{N_{TSA}/T_{TSA}}{N_{input}/T_{input}}\right) \quad (1)$$

SON TSA-Seq correlations with other genomic features
We partitioned the genome into nonoverlapping 20-kb bins, averaging the TSA-Seq score (Eq. 1) over the 100-bp intervals within this bin for the new TSA-Seq 20-kb score. The values for these 20-kb bins were then smoothed by convolution using a Hanning window of length 21 (21 × 20 kb or 420 kb). This new smoothed TSA-Seq score with 20-kb sampling was used for SON TSA-Seq decile calculations and comparisons with other genome features.

Various functional genomic datasets available for K562 cells were downloaded from the ENCODE project portal (https://www.encodeproject.org/; Table S4). For gene expression analysis, we used mRNA fragments per kilobase of transcript per million

mapped reads (FPKM) values based on RNA-Seq data for protein-coding genes. Gene density was calculated by taking the fraction of protein-coding genes in each SON TSA-Seq decile. For DNase I–hypersensitive sites, we calculated the number of DNase-Seq peaks contained within each decile. Histone modification and transcription factor ChIP-Seq data are classified into punctuated peaks, broad peaks, or a mixture of punctuated and broad peaks based on signal enrichment patterns. ChIP-Seq data (H3K27ac, H3K4me1, H3K4me2, H3K4me3, H3K79me2, H3K9ac, and CTCF) with punctuated peak patterns show sharply defined peaks of localized enrichment. ChIP-Seq data (H2A.Z, H3K27me3, H3K36me3, H3K9me1, H3K9m3, and RNA pol II) with broad peak patterns show diffuse widespread genomic ChIP-Seq signals without discrete peaks. ChIP-Seq data with a mixture pattern of punctuated peaks and broad peak show both local and large-scale signal enrichment (H4K20me1).

For ChIP-Seq data with punctuated peaks, we counted peak number for each SON TSA-Seq decile. For ChIP-Seq data with broad peaks, we processed the ChIP-Seq data using a similar sliding window method as used for the TSA-Seq data (Eq. 2) using normalized ENCODE bigWig files for both the ChIP sample and input DNA. N_{ChIP} is the sum of signals in a 20-kb window for the ChIP sample. N_{input} is the sum of signals in a 20-kb window for the input DNA. EN_{input} is the expected value for a 20-kb window based on the total number of mapped reads and the total genome-mappable base pairs. N_{ChIP_norm} can be interpreted as the ChIP signal normalized by the input DNA signal. The mode of the normalized ChIP scores, N_{ChIP_norm}, was considered as the nonspecific antibody-binding background level, which was subtracted from the normalized score in each 20-kb window to calculate the final ChIP-Seq enrichment score, E_{ChIP} (Eq. 3). The summed ChIP-Seq enrichment scores were calculated for each decile. An exception was H3K9me3, for which real signals were present throughout the genome. For this mark, the 0.5% percentile value was assumed as background and subtracted from the normalized score. For data with a mixture of sharp and broad peaks, we calculated the total number of base pairs within the ChIP-Seq peaks (peak length) in each SON TSA decile.

Genome guanine-cytosine (GC) content data were downloaded from the UCSC Genome Browser, and the GC fraction in each 20-kb bin was calculated for each decile. We downloaded the RepeatMasker track for hg19 from the UCSC Genome Browser to calculate distributions of different repeat types in each decile. We calculated these distributions using coverage for long interspersed nuclear element (LINE) repeats but repeat count for the shorter Alu, short interspersed nuclear element (SINE), simple, and satellite repeats.

$$N_{ChIP\_norm} = \frac{N_{ChIP}}{N_{input}/EN_{input}} \quad (2)$$

$$E_{ChIP} = \max\left(0, N_{ChIP\_norm} - Mode_{N_{ChIP_{norm}}}\right) \quad (3)$$

Replication timing and Hi-C subcompartment data from K562 and GM12878 cells, respectively, were compared with SON TSA-Seq deciles in two ways: (1) calculating the fraction of each replication timing domain or Hi-C subcompartment group overlapping each SON TSA-Seq decile, and (2) calculating the

fraction of each SON TSA-Seq decile that was comprised of each replication timing domain or each Hi-C subcompartment group. For display of replication timing, we also used the University of Washington ENCODE Repli-Seq track corresponding with the wavelet-smoothed signal that combines data from different replication timing domains into one track.

For analysis of LAD distribution among the SON TSA-Seq deciles, we first followed the algorithm described previously for LAD segmentation (Guelen et al., 2008). Adjacent LADs were then merged if >80% of probes of the region between two LADs had positive log₂ ratios. For analysis of distribution of enhancers and super-enhancers across SON TSA-Seq deciles, we downloaded genomic locations of K562 super-enhancers from Hnisz et al. (2013). The SON TSA decile assignment for each enhancer was picked by whichever had the largest fraction over the enhancer region.

Box plots
Each box plot shows median (inside line) and 25th (box bottom) and 75th percentiles (box top). The top whisker extends from the 75th percentile to the highest value within 1.5-fold of the box length, and the bottom whisker extends from the 25th percentile to the lowest value within 1.5-fold of the box length. Outliers are marked as black dots. Means are marked with red diamonds if shown.

Quality control with TSA-Seq
TSA-Seq is fundamentally different from molecular proximity methods such as ChIP-Seq. The amount of biotin-DNA labeling produced by the TSA reaction at a given chromosome locus is proportional to the intensity of the tyramide-biotin signal over that chromosome locus. Therefore, the most essential control for TSA staining is direct microscopic visualization of the biotin intranuclear distribution after TSA staining. Labeling specificity should be validated by comparing the biotin staining with traditional immunofluorescence staining using the same primary antibody. TSA labeling should show both a specific labeling with a high intensity at the target compartment as well as a diffuse signal surrounding the target, which decays with distance from the target. We aliquoted cells for biotin staining and microscopic inspection after each batch of TSA staining. Batches showing unsatisfactory TSA staining were discarded. This was followed by dot blot analysis of biotin DNA labeling both before and after affinity purification. Again, samples that did not show satisfactory biotin enrichment over negative control (no primary antibody) samples were discarded.

TSA-Seq is different because the tyramide free radical spreads over hundreds of nanometers to a micrometer distance away from the staining target. ChIP-Seq results may vary between antibodies probing different components of a multisubunit protein complex because of their varying distances to DNA and ability to cross-link to DNA. The target proteins themselves may show varying ability to cross-link to chromatin and survive the chromatin solubilization step of ChIP-Seq. In contrast, antibodies targeting different subunits of protein complexes should produce similar TSA-Seq results as long as the distance between these target subunits is small relative to the ~10–100-nm distance

over which the concentration of tyramide free radical noticeably decays. Even antibodies targeting different protein complexes should produce similar TSA-Seq results as long as these different complexes are colocalized over a distance that is small relative to the microscopically observed distance over which the tyramide free radical concentration noticeably decays. For example, for a decay constant of –4 (condition 2), the TSA signal would drop only 4% at a distance of 10 nm from the target, 18% at a distance of 50 nm, and 33% at a distance of 100 nm.

Conversely, even ChIP-Seq–validated antibodies may perform poorly with TSA-Seq. An antibody may show strong nonspecific or off-target binding without affecting ChIP results if the antibody is too far to effectively cross-link to DNA. Yet this nonspecifically or off-target bound antibody will produce a tyramide free radical signal that will diffuse and react with neighboring DNA, and the amount of DNA reaction will be linearly proportional to the intensity of this nonspecific biotin signal as visualized by light microscopy.

Generation of polyclonal antibody against SON
Rabbit polyclonal antibody against hSON peptide (947–964; Cys-DPYRLGHDPYRLTPDPYR) was produced by Pacific Immunology Corp. Antigen peptide was synthesized, conjugated to the carrier protein keyhole limpet hemocyanin, and affinity purified for immunization. Antibody was affinity purified against the antigen peptide from the third bleed antiserum and stored in 50% glycerol solution at a concentration of 1.6 mg/ml.

Generation of monoclonal antibodies against lamin A/C and lamin B
All of the ROD2 and tail domains of lamin A/C and lamin B (LB1 and LB2) were expressed in bacteria as GST fusion proteins. They were purified by FPLC and then mixed together before injection into mice for immunization. Western blotting was used to screen the monoclonal antibodies from hybridoma colonies for their lamin specificity. Monoclonal antibody from clone 5G4 showed lamin A/C staining with no cross-reactivity apparent at a dilution of 1:10,000 and only trace reactivity at a dilution of 1:500. Monoclonal antibody from clone 2D8 showed lamin B staining with no cross-reactivity apparent at a dilution of 1:2,000 and only trace reactivity to lamin A/C at lower dilutions. Dilutions of monoclonal antibodies for TSA staining were of concentrated Bioreactor supernatants.

Structured illumination microscopy (SIM)
3D SIM images were taken with 0.125-μm z steps using an OMX Blaze V4 (GE Healthcare) with a 100×, 1.4 NA oil immersion objective lens (Olympus), laser illumination (405, 488, and 568 nm), and an Evolve 512 electron-multiplying charge-coupled device camera (Photometrics). Solid-state illumination was used for Cy5 in widefield mode for the fourth channel (SON) in Fig. 5 C. Image reconstruction, channel registration, and alignment were performed with SoftWoRx.

Bioinformatics analysis of other genome features
For correlating SON TSA-Seq levels relative to DNA repeats, we recorded positions for different repeat classes/families using the

Chen et al.
TSA-Seq mapping of nuclear genome organization

Journal of Cell Biology
https://doi.org/10.1083/jcb.201807108
4044

UCSC RepeatMasker track. The GC content in each SON TSA-Seq decile was calculated using the nucBed program included in bedtools (version 2.17; Quinlan and Hall, 2010). The number of exons, introns, isoforms, and transcription start sites (TSS) were extracted from the GENCODE (Harrow et al., 2012) annotation file version 19 from the GENCODE website (http://www.gencodegenes.org/). Gene sizes and gene density across SON TSA-Seq deciles were extracted from the GENCODE annotation file version 3c, the same annotation file used for generating gene expression data.

Hi-C compartment eigenvector

To generate the Hi-C compartment eigenvector track, we used a recently published high resolution Hi-C dataset (Rao et al., 2014) from human K562 cells. First, we downloaded the intra-chromosome Hi-C raw contact matrices at 10-kb resolution and normalized the contact matrices using provided KR-normalization (Knight and Ruiz, 2013) vectors. Then we applied principal component analysis using the function "princomp" with default parameters in R on the normalized contact matrix of each chromosome individually. As shown previously (Imakaev et al., 2012), eigenvector expansion is equivalent to principal component analysis of the normalized contact matrix. The first principle component (PC1) corresponds with the eigenvector with the largest eigenvalue. Eigenvectors were normalized by magnitude. In this study, we only showed the first eigenvector.

Applying the normalization method used for TSA-Seq to Pol II ChIP-Seq data

The normalization scheme used for TSA-Seq data, which computes relative enrichment/depletion over a certain bin size, can also be applied to ChIP-Seq data. We processed the ENCODE K562 Pol II ChIP-Seq and input raw fastq reads using the same alignment pipeline as used for the SON TSA-Seq data. Alignments from replicate ChIP-Seq experiments were merged. We calculated the normalized TSA-Seq enrichment score using Eq. 4. In this method, N_{Pol2} is the number of mapped reads in a 20-kb window in the RNA Pol II ChIP-Seq sample (N_{input} is defined similarly for input data). B_{Pol2} is the expected number of reads from the background-nonspecific binding of RNA Pol II antibody to chromatin. To estimate B_{Pol2}, we first ranked the genomic intergenic regions from longest to shortest. We used the top 5% longest intergenic regions, which we assumed to be gene deserts, to estimate background reads. For these largest intergenic regions, we calculated the average read count per base pair in the ChIP-Seq data. B_{Pol2} is the product of the size in base pairs of the sliding window with this average background read count per base pair. The background adjustment for each bin is done by using the max (0, $N_{Pol2} - B_{Pol2}$). Finally, T_{Pol2} is the total number of mapped reads for ChIP-Seq sample after the background subtraction for each bin, and T_{input} is the total number of mapped reads in input data.

$$\text{Pol II ChIP-Seq enrichment score} =$$
$$\log_2 \left(\frac{\max(0, \, N_{Pol2} - B_{Pol2})/T_{Pol2}}{N_{input}/T_{input}} \right) \qquad (4)$$

Transcription hot zones

To identify transcription hot zones, we started with the smoothed RNA Pol II ChIP-Seq track. We identified all genomic regions with positive values >0.1, and for each of these regions, we matched it to the local maximum in the SON TSA-Seq score within this same region. If more than one local maximum was present within the region, we used the largest local maximum. In nearly all cases, these SON TSA-Seq local maxima aligned with the local maxima in the smoothed RNA Pol II track. We identified the Hi-C subcompartment assignment corresponding to these SON TSA-Seq local maxima as well as their SON TSA-Seq scores and corresponding percentiles. We then calculated the distance distribution for the A1 and A2 SON TSA-Seq local maxima to the nearest LAD/inter-LAD boundary.

2D TSA-Seq scatter plots

For 2D SON or lamin B TSA-Seq and 1-Mbp smoothed normalized RNA Pol II ChIP-Seq scatter plots (Fig. S3 C), each point represents a 160-kb bin for which the TSA-Seq average score was calculated from the smoothed TSA-Seq score. These points were plotted versus score percentile.

We also used similar 2D scatter plots to show the relationship of SON, pSC35, lamin A/C, lamin B, and Pol II TSA-Seq data in Fig. 5 B. Each point represents a 300-kb bin from the smoothed TSA-Seq score. These points were plotted versus score value or percentile. We also used these 2D scatter plots to show the relationship between lamin and SON TSA-Seq with gene expression or DNA replication timing. In these scatter plots (Fig. 8, E–G; and Fig. S5), the x–y axes correspond with either TSA-Seq score values or percentiles of the TSA-Seq score over the entire genome, with distribution projections shown on the sides.

For gene expression, we calculated the FPKM of each gene and divided genes into two groups: expressed (FPKM > 0) and nonexpressed (FPKM = 0). For expressed genes, we further separated them into 10 deciles, ranking them based on their FPKM values. Each dot represents a single gene. For DNA replication timing, we used the UW Repli-Seq data for K562 cells (Hansen et al., 2010). Each dot represents a 300-kb genome bin. The fraction of each 300-kb bin replicating in each of the measured replication timing domains (G1, S1, S2, S3, and S4) was calculated, and the bin was assigned to that replication domain with the largest fraction.

For computing the 2D TSA-Seq scatter plot showing the distribution of Hi-C subcompartments (Fig. 5 F), we computed the average lamin B or SON TSA-Seq scores in 320-kb nonoverlapping bins calculated from the smoothed 20-kb bin TSA-Seq scores as described above for the other 2D scatter plots. For each 320-kb bin, the fraction of this bin corresponding with each Hi-C subcompartment was calculated. The Hi-C subcompartment with the largest fraction was then assigned to this 320-kb bin. TSA-Seq scores were plotted as percentiles.

2D TSA-Seq plots of weighted, wavelet-smoothed DNA replication timing

We again used the UW replication timing data for K562 cells but now used the weighted, wavelet-smoothed composite signal that measures relative timing of DNA replication such that earlier replicating regions have higher values and are ranked as percentiles.

Average replication timing scores were calculated for 160-kb genomic bins. In Fig. 8 G, each 160-kb genomic bin was mapped to a given lamin B and SON TSA-Seq percentile based on the average TSA-Seq score calculated for each 160-kb bin using the smoothed TSA-Seq 20-kb bin data. We then divided the lamin B and SON TSA-Seq percentile ranges into hexagonal pixels spaced by two percentiles. We calculated the median replication timing of all 160-kb genomic bins mapping to a given hexagon. These median values of replication percentiles were color-coded and plotted.

Gene density near type I and type II peaks
We separated ±500 kb of transcription hot zones into five groups based on their distance to peak summit (~0–100 kb, ~100–200 kb, ~200–300 kb, ~300–400 kb, and ~400–500 kb). Within each group, we counted the number of overlapping protein-coding genes. The gene density was calculated as the average number of genes per 100 kb. Besides the gene density for all genes, we also calculated the gene density for specific groups of genes based on expression level (top 10% expressed genes, ~40–60% expressed genes, and bottom 10% expressed genes).

Pausing index of genes in TSA-Seq peaks
We calculated the pausing index of each gene based on Pol II ChIP-Seq. The Pol II ChIP-Seq was downloaded from ENCODE (ENCSR000FAJ). The method of calculating pausing index was described by Day et al. (2016).

$$Pausing\ Index = \frac{ReadCount(TSSR)/Length(TSSR)}{ReadCount(Gene\ Body)/Length(Gene\ Body)}$$

TSS region, –50 to +300 bp around TSS; gene body, TSS + 300 bp to +3 kb past the annotated transcriptional end site.

Then, we identified genes that were within the ±100-kb range of TSA-Seq peaks. TSA-Seq peaks were separated into two groups based on whether they were in A1 or A2 subcompartments. In Fig. 10 B, we plotted the density of the pausing index of genes near (±200 kb) A1/A2 TSA-Seq peaks.

Housekeeping genes in TSA-Seq deciles
A list of 3,904 human housekeeping genes was downloaded from http://www.tau.ac.il/~elieis/HKG/HK_genes.txt, which is an updated version of the original release from Eisenberg and Levanon (2013). To determine the nonhousekeeping genes as well as the genomic position of these genes, we extracted protein-coding genes with completed CDS annotation from the hg19 RefSeq gene annotation (downloaded from the UCSC Genome Browser). 3,791 genes among the original 3,904 housekeeping genes were found to have matched gene ID or gene name in the RefSeq gene annotation. Therefore, the remaining 15,751 protein-coding genes were determined as nonhousekeeping genes. To compare the distribution of housekeeping genes to nonhousekeeping genes across a combination of TSA-Seq decile and Hi-C subcompartment, we determined the number of genes overlapping with a specific SON TSA-Seq decile (e.g., decile 1) within a particular Hi-C subcompartment (e.g., A1). The number of overlapped genes were further normalized to a percentage for each type of Hi-C subcompartment.

Data availability
All high throughput sequencing data generated by TSA-Seq and microarray data generated by DamID are deposited at GEO under GSE66019.

Code availability
All computational code used to analyze TSA-Seq data are available at https://github.com/ma-compbio/TSA-Seq-toolkit.

Online supplemental material
Fig. S1 shows TSA-Seq measures cytological proximity. Fig. S2 shows 3D immuno-FISH validation that SON TSA-Seq score measures cytological mean distance to nuclear speckles. Fig. S3 shows TSA-Seq maps predict both chromosome location and chromosome topology. Fig. S4 shows genome features relative to SON TSA-Seq decile. Fig. S5 shows 2D scatter plots showing distribution of gene expression and DNA replication timing relative to speckle–lamin TSA-Seq axis. Table S1 shows a summary of conditions for all TSA-Seq experiments. Table S2 shows BACs used for DNA FISH. Table S3 shows a list of protein-coding genes and their SON TSA-Seq deciles and expression levels. Table S4 shows the public datasets analyzed.

References

Agard, D.A., Y. Hiraoka, P. Shaw, and J.W. Sedat. 1989. Fluorescence microscopy in three dimensions. *Methods Cell Biol.* 30:353–377. https://doi.org/10.1016/S0091-679X(08)60986-3

Belmont, A.S., and K. Bruce. 1994. Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J. Cell Biol.* 127:287–302. https://doi.org/10.1083/jcb.127.2.287

Bickmore, W.A. 2013. The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* 14:67–84. https://doi.org/10.1146/annurev-genom-091212-153515

Bobrow, M.N., T.D. Harris, K.J. Shaughnessy, and G.J. Litt. 1989. Catalyzed reporter deposition, a novel method of signal amplification. Application to immunoassays. *J. Immunol. Methods.* 125:279–285. https://doi.org/10.1016/0022-1759(89)90104-X

Bobrow, M.N., K.J. Shaughnessy, and G.J. Litt. 1991. Catalyzed reporter deposition, a novel method of signal amplification. II. Application to membrane immunoassays. *J. Immunol. Methods.* 137:103–112. https://doi.org/10.1016/0022-1759(91)90399-Z

Brangwynne, C.P. 2011. Soft active aggregates: mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies. *Soft Matter.* 7:3052–3059. https://doi.org/10.1039/c0sm00981d

Carter, D., L. Chakalova, C.S. Osborne, Y.F. Dai, and P. Fraser. 2002. Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* 32:623–626. https://doi.org/10.1038/ng1051

Carter, K.C., K.L. Taneja, and J.B. Lawrence. 1991. Discrete nuclear domains of poly(A) RNA and their relationship to the functional organization of the nucleus. *J. Cell Biol.* 115:1191–1202. https://doi.org/10.1083/jcb.115.5.1191

Ching, R.W., K. Ahmed, P.C. Boutros, L.Z. Penn, and D.P. Bazett-Jones. 2013. Identifying gene locus associations with promyelocytic leukemia nuclear bodies using immuno-TRAP. *J. Cell Biol.* 201:325–335. https://doi.org/10.1083/jcb.201211097

Day, D.S., B. Zhang, S.M. Stevens, F. Ferrari, E.N. Larschan, P.J. Park, and W.T. Pu. 2016. Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types. *Genome Biol.* 17:120. https://doi.org/10.1186/s13059-016-0984-2

Dernburg, A.F. 2011. Fragmentation and labeling of probe DNA for whole-mount FISH in Drosophila. *Cold Spring Harb. Protoc.* 2011:1527–1530. https://doi.org/10.1101/pdb.prot066886

Dow, E.C., H. Liu, and A.P. Rice. 2010. T-loop phosphorylated Cdk9 localizes to nuclear speckle domains which may serve as sites of active P-TEFb function and exchange between the Brd4 and 7SK/HEXIM1 regulatory complexes. *J. Cell. Physiol.* 224:84–93. https://doi.org/10.1002/jcp.22096

Eisenberg, E., and E.Y. Levanon. 2013. Human housekeeping genes, revisited. *Trends Genet.* 29:569–574. https://doi.org/10.1016/j.tig.2013.05.010

Fakan, S., and E. Puvion. 1980. The ultrastructural visualization of nucleolar and extranucleolar RNA synthesis and distribution. *Int. Rev. Cytol.* 65:255–299. https://doi.org/10.1016/S0074-7696(08)61962-2

Galganski, L., M.O. Urbanek, and W.J. Krzyzosiak. 2017. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res.* 45:10350–10368. https://doi.org/10.1093/nar/gkx759

Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M.B. Faza, W. Talhout, B.H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature.* 453:948–951. https://doi.org/10.1038/nature06947

Hall, L.L., K.P. Smith, M. Byron, and J.B. Lawrence. 2006. Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 288:664–675. https://doi.org/10.1002/ar.a.20336

Hansen, R.S., S. Thomas, R. Sandstrom, T.K. Canfield, R.E. Thurman, M. Weaver, M.O. Dorschner, S.M. Gartler, and J.A. Stamatoyannopoulos. 2010. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. USA.* 107:139–144. https://doi.org/10.1073/pnas.0912402107

Harrow, J., A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B.L. Aken, D. Barrell, A. Zadissa, S. Searle, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22:1760–1774. https://doi.org/10.1101/gr.135350.111

Hendzel, M.J., M.J. Kruhlak, and D.P. Bazett-Jones. 1998. Organization of highly acetylated chromatin around sites of heterogeneous nuclear RNA accumulation. *Mol. Biol. Cell.* 9:2491–2507. https://doi.org/10.1091/mbc.9.9.2491

Herrmann, C.H., and M.A. Mancini. 2001. The Cdk9 and cyclin T subunits of TAK/P-TEFb localize to splicing factor-rich nuclear speckle regions. *J. Cell Sci.* 114:1491–1503.

Hnisz, D., B.J. Abraham, T.I. Lee, A. Lau, V. Saint-André, A.A. Sigova, H.A. Hoke, and R.A. Young. 2013. Super-enhancers in the control of cell identity and disease. *Cell.* 155:934–947. https://doi.org/10.1016/j.cell.2013.09.053

Hopman, A.H.N., F.C.S. Ramaekers, and E.J.M. Speel. 1998. Rapid synthesis of biotin-, digoxigenin-, trinitrophenyl-, and fluorochrome-labeled tyramides and their application for In situ hybridization using CARD amplification. *J. Histochem. Cytochem.* 46:771–777. https://doi.org/10.1177/002215549804600611

Hu, Y., I. Kireev, M. Plutz, N. Ashourian, and A.S. Belmont. 2009. Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. *J. Cell Biol.* 185:87–100. https://doi.org/10.1083/jcb.200809196

Imakaev, M., G. Fudenberg, R.P. McCord, N. Naumova, A. Goloborodko, B.R. Lajoie, J. Dekker, and L.A. Mirny. 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods.* 9:999–1003. https://doi.org/10.1038/nmeth.2148

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

Khanna, N., Y. Hu, and A.S. Belmont. 2014. HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.* 24:1138–1144. https://doi.org/10.1016/j.cub.2014.03.053

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

Knight, P.A., and D. Ruiz. 2013. A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* 33:1029–1047. https://doi.org/10.1093/imanum/drs019

Ko, T.K., E. Kelly, and J. Pines. 2001. CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J. Cell Sci.* 114:2591–2603.

Kölbl, A.C., D. Weigl, M. Mulaw, T. Thormeyer, S.K. Bohlander, T. Cremer, and S. Dietzel. 2012. The radial nuclear positioning of genes correlates with features of megabase-sized chromatin domains. *Chromosome Res.* 20:735–752. https://doi.org/10.1007/s10577-012-9309-9

Landt, S.G., G.K. Marinov, A. Kundaje, P. Kheradpour, F. Pauli, S. Batzoglou, B.E. Bernstein, P. Bickel, J.B. Brown, P. Cayting, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22:1813–1831. https://doi.org/10.1101/gr.136184.111

Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9:357–359. https://doi.org/10.1038/nmeth.1923

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lieberman-Aiden, E., N.L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 326:289–293. https://doi.org/10.1126/science.1181369

Meuleman, W., D. Peric-Hupkes, J. Kind, J.B. Beaudry, L. Pagie, M. Kellis, M. Reinders, L. Wessels, and B. van Steensel. 2013. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* 23:270–280. https://doi.org/10.1101/gr.141028.112

Moen, P.T. Jr., C.V. Johnson, M. Byron, L.S. Shopland, I.L. de la Serna, A.N. Imbalzano, and J.B. Lawrence. 2004. Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol. Biol. Cell.* 15:197–206. https://doi.org/10.1091/mbc.e03-06-0388

Quinlan, A.R., and I.M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 26:841–842. https://doi.org/10.1093/bioinformatics/btq033

Rao, S.S.P., M.H. Huntley, N.C. Durand, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, A.L. Sanborn, I. Machol, A.D. Omer, E.S. Lander, and E.L. Aiden. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 159:1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods.* 9:676–682. https://doi.org/10.1038/nmeth.2019

Shopland, L.S., C.V. Johnson, M. Byron, J. McNeil, and J.B. Lawrence. 2003. Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains. *J. Cell Biol.* 162:981–990. https://doi.org/10.1083/jcb.200303131

Sims, J.K., S.I. Houston, T. Magazinnik, and J.C. Rice. 2006. A trans-tail histone code defined by monomethylated H4 Lys-20 and H3 Lys-9 demarcates distinct regions of silent chromatin. *J. Biol. Chem.* 281:12760–12766. https://doi.org/10.1074/jbc.M513462200

Solovei, I., and M. Cremer. 2010. 3D-FISH on cultured cells combined with immunostaining. *Methods Mol. Biol.* 659:117–126. https://doi.org/10.1007/978-1-60761-789-1_8

Spector, D.L., and A.I. Lamond. 2011. Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* 3:a000646. https://doi.org/10.1101/cshperspect.a000646

Takizawa, T., K.J. Meaburn, and T. Misteli. 2008. The meaning of gene positioning. *Cell.* 135:9–13. https://doi.org/10.1016/j.cell.2008.09.026

van Gijlswijk, R.P., J. Wiegant, A.K. Raap, and H.J. Tanke. 1996. Improved localization of fluorescent tyramides for fluorescence in situ hybridization using dextran sulfate and polyvinyl alcohol. *J. Histochem. Cytochem.* 44:389–392. https://doi.org/10.1177/44.4.8601698

van Gijlswijk, R.P.M., H.J.M.A.A. Zijlmans, J. Wiegant, M.N. Bobrow, T.J. Erickson, K.E. Adler, H.J. Tanke, and A.K. Raap. 1997. Fluorochrome-labeled tyramides: use in immunocytochemistry and fluorescence in situ hybridization. *J. Histochem. Cytochem.* 45:375–382. https://doi.org/10.1177/002215549704500305

van Steensel, B., and A.S. Belmont. 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell.* 169:780–791. https://doi.org/10.1016/j.cell.2017.04.022

Vogel, M.J., D. Peric-Hupkes, and B. van Steensel. 2007. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.* 2:1467–1478. https://doi.org/10.1038/nprot.2007.148

Wang, G., C.L. Achim, R.L. Hamilton, C.A. Wiley, and V. Soontornniyomkij. 1999. Tyramide signal amplification method in multiple-label immunofluorescence confocal microscopy. *Methods.* 18:459–464. https://doi.org/10.1006/meth.1999.0813

Wang, Z., C. Zang, J.A. Rosenfeld, D.E. Schones, A. Barski, S. Cuddapah, K. Cui, T.-Y. Roh, W. Peng, M.Q. Zhang, and K. Zhao. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40:897–903. https://doi.org/10.1038/ng.154

Wartlick, O., A. Kicheva, and M. González-Gaitán. 2009. Morphogen gradient formation. *Cold Spring Harb. Perspect. Biol.* 1:a001255. https://doi.org/10.1101/cshperspect.a001255

Xie, S.Q., S. Martin, P.V. Guillot, D.L. Bentley, and A. Pombo. 2006. Splicing speckles are not reservoirs of RNA polymerase II, but contain an inactive form, phosphorylated on serine2 residues of the C-terminal domain. *Mol. Biol. Cell.* 17:1723–1733. https://doi.org/10.1091/mbc.e05-08-0726

Xing, Y., C.V. Johnson, P.T. Moen Jr., J.A. McNeil, and J. Lawrence. 1995. Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J. Cell Biol.* 131:1635–1647. https://doi.org/10.1083/jcb.131.6.1635

Zhu, L., and C.P. Brangwynne. 2015. Nuclear bodies: the emerging biophysics of nucleoplasmic phases. *Curr. Opin. Cell Biol.* 34:23–30. https://doi.org/10.1016/j.ceb.2015.04.003

Zweidler, A. 1992. Role of individual histone tyrosines in the formation of the nucleosome complex. *Biochemistry.* 31:9205–9211. https://doi.org/10.1021/bi00153a013