SPATIAL ORGANIZATION OF TRANSCRIBED EUKARYOTIC GENES

Susanne Leidescher[1], Johannes Nübler[2], Yana Feodorova[1,3], Erica Hildebrand[4], Simon Ullrich[1], Sebastian Bultmann[1], Stephanie Link[5], Katharina Thanisch[1,6], Job Dekker[4,7], Heinrich Leonhardt[1], Leonid Mirny[2], Irina Solovei[1]*

[1] Department of Biology II, Biozentrum, Ludwig-Maximilians University Munich (LMU), Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany

[2] Institute for Medical Engineering and Science, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3] Department of Medical Biology, Medical University of Plovdiv; Division of Molecular and Regenerative Medicine, Research Institute at Medical University of Plovdiv, Boulevard Vasil Aprilov 15A, Plovdiv 4000, Bulgaria

[4] Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

[5] BioMedizinisches Center, Ludwig-Maximilians University Munich, Grosshaderner Str. 9, 82152 Planegg-Martinsried, Germany

[6] *Current address:* Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Str. 65, 88397, Biberach an der Riss, Germany

[7] Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

* Correspondence: Irina.Solovei@lrz.uni-muenchen.de

SUMMARY

Despite the well established role of nuclear organization in the regulation of gene expression, little is known about the reverse: how transcription shapes the spatial organization of the genome. In particular, given the relatively small sizes of genes and the limited resolution of light microscopy, the structure and spatial arrangement of a single transcribed gene are still poorly understood. Here, we make use of several long highly expressed mammalian genes and demonstrate that they form Transcription Loops (TLs) with polymerases moving along the loops and carrying nascent RNAs that undergo co-transcriptional splicing. TLs dynamically modify their harboring loci and extend into the nuclear interior suggesting an intrinsic stiffness. Both experimental evidence and polymer modeling support the hypothesis that TL stiffness is caused by the dense decoration of transcribed genes with multiple voluminous nascent RNPs. We propose that TL formation is a universal principle of eukaryotic gene expression.

INTRODUCTION

The understanding of eukaryotic gene transcription and the mechanisms of its regulation is progressively increasing at both the molecular level and at the level of the cell nucleus. At the molecular level, our knowledge of transcription comprises the fine mechanisms of enhancer and promoter dynamics (Hnisz et al., 2017; Schoenfelder and Fraser, 2019), transcription initiation, elongation and termination (Liu et al., 2013; Proudfoot, 2016), the RNA polymerase II (RNAPII) transcription cycle (Cramer, 2019; Eick and Geyer, 2013), co-transcriptional pre-mRNA processing (Herzel et al., 2018; Nojima et al., 2018), remodeling of chromatin during transcription (Becker and Workman, 2013), and many other aspects of transcription regulation. On a more global scale, studies of nuclear architecture revealed the importance of spatial segregation of active euchromatin from inactive heterochromatin: silenced heterochromatin resides in the repressive compartment of the nuclear periphery, whereas transcriptionally active euchromatin occupies the nuclear interior enriched in transcription and splicing machineries (Feodorova et al., 2020; Solovei et al., 2016; van Steensel and Belmont, 2017).

Knowledge pertaining to an intermediate level of transcription organization, i.e. the spatial arrangement of a single expressed gene, is however surprisingly limited and essentially represented by the hypothesis about so-called *transcription factories*. The hypothesis postulates that RNAPIIs are immobilized in groups, and upon transcription activation, genes from the same or different chromosomes approach a transcription factory and then are reeled through immobilized RNAPIIs with nascent RNAs (nRNAs) extruded in a spot (Cook, 1999; Papantonis and Cook, 2013). Indeed, detection of elongating RNAPIIs and nRNAs, at both global and single gene levels (Brown et al., 2008; Osborne et al., 2004; Schoenfelder et al., 2010) reveals foci-like signals, considered as evidence for the existence of transcription factories (Chakalova and Fraser, 2010). Immobility of RNAPIIs in transcription factories is, however, at odds with observations on transcription units with fixed ends, such as the lateral loops of lampbrush chromosomes (Macgregor, 1993), Y-loops of *Drosophila* spermatocytes (Grond et al., 1983) and puffs of polytene chromosomes (Bjork and Wieslander, 2015). In these cases, RNAPIIs have been shown moving along a transcription unit and carrying a cargo of progressively growing nRNAs.

The above discrepancy can be explained by the different features of the studied genes. Transcription units of lampbrushes exceed gene length, include repetitive sequences that add up to hundreds of kilobases, and measure from several to hundreds of microns (Macgregor, 1993). Importantly, in lateral loops of lampbrushes, Y-loops and puffs, such as Balbiani rings, the level of transcription is so high that, due to dense decoration by thousands of RNAPIIs with nRNAs, transcribed regions gain a noticeable width. This in combination with substantial length, allows their visualization even under a phase contrast microscope (Bjork and Wieslander, 2015; Grond et al., 1983; Morgan, 2018). Therefore, for studying expressed genes by light microscopy, both sufficient length and sufficiently high expression are essential prerequisites. The combination of these two traits, however, is rarely met in cultured mammalian cells that are the major source of knowledge about transcription. Indeed, the majority of studied highly expressed genes are short (e.g., *HBB* is only 1.6 kb) and given that a 10 kb-long gene, when fully stretched, measures only 0.5 µm, it is understandable why the chance to resolve it microscopically is close to zero. At the same time, long genes are in general not highly expressed and especially not in cultured cells (Jjingo et al., 2011) (Human Protein Atlas 19.3).

To fill the gap in knowledge about the spatial organization of expressed genes, we selected several genes that are both relatively long and relatively highly expressed and studied their spatial arrangement in differentiated mouse cells. We demonstrate that these genes form microscopically resolvable transcription loops (TLs) similar to lampbrush loops and polytene puffs. We provide evidence that TLs are decorated by elongating RNAPIIs moving along the gene axis and carrying nRNAs that undergo co-transcriptional splicing. Furthermore, we show that TLs dynamically modify their harboring loci and extend into the nuclear interior. We hypothesize that the extension of TLs results from an increased stiffness of highly expressed long genes due to their decoration by voluminous nascent RNPs and provide evidence supporting this hypothesis, including polymer modeling of TLs. In conclusion, we suggest that TL formation is a universal principle of eukaryotic gene transcription.

RESULTS

Long mammalian genes are rarely highly expressed

For the microscopic visualization of transcribed genes, we searched for genes that are both relatively long and highly expressed, using arbitrary thresholds for length such as ≥100 kb, corresponding to the size of the smallest discernible lampbrush loops (Hutchison, 1987), and for expression level such as ≥ 1,000 TPM (transcripts per million), corresponding to the average expression level of the human GAPDH gene (GTEx Consortium; https://www.gtexportal.org/home/).

Analysis of gene length and expression levels within human and mouse genomes showed that less than 20% of genes are ≥100 kb (Fig. S1A) and that such genes, as a rule, are lowly expressed (Fig. S1B-D). Based on GTEx RNA-seq data, we selected 10 human genes that were above our thresholds and out of these further selected only protein coding genes that are expressed in cell types unambiguously identifiable in tissue sections (SI Table S1). The most highly expressed gene among them was the thyroglobulin gene (*TG*) coding for the extracellular protein thyroglobulin secreted by thyroid gland cells (Fig.1A). The other four selected genes encoded structural proteins of the contractile machineries of skeletal and smooth muscles (Fig.1A).

Next, we confirmed the high expression of the selected five genes in the corresponding mouse tissues (Fig.1A; SI Table S2). RNA-seq analysis of thyroid gland revealed that the mouse thyroglobulin gene (*Tg*) with a length of 180 kb is exceptionally upregulated (22,924 TPM) with an expression level 10-fold higher than ubiquitously and strongly expressed housekeeping genes such as non-muscle actin (*Actb*, 2,791 TPM) or ribosomal protein L41 (*Rpl41*, 2,467 TPM). RNA-seq analysis of bladder tissue enriched in smooth muscles confirmed high levels of expression of *Myh11* (97 kb) and *Cald1* (177 kb) at 2,179 TPM and 1,698 TPM, respectively. For RNA-seq analysis of skeletal muscle we chose to use myotubes differentiated from myoblasts *in vitro* and measured the expression level of *Ttn* (279 kb) and *Neb* (242 kb) at 2,100 TPM and at 980 TPM, respectively. In addition, we noticed that *Cald1* is relatively highly expressed in myoblasts (1,797 TPM) (Fig.1A; SI Table S2). Thus, the five selected mouse genes met the conditions we have defined as prerequisites for a study of transcribed genes by light microscopy.

Highly expressed genes form Transcription Loops

We reasoned that the considerable length of these genes should enable their microscopical visualization similar to the visualization of lampbrush chromosomes' lateral loops and puffing genes of polytene chromosomes. Therefore, we carried out DNA-FISH using genomic BAC probes encompassing the selected genes (SI Table S3) in cryosections of the corresponding mouse tissues – thyroid gland (for *Tg*), skeletal muscle (for *Ttn* and *Neb*), heart muscle (for *Ttn*), bladder and colon (for *Myh11* and *Cald1*). In case of *Ttn, Neb* and *Cald1*, FISH was also performed on cultured myoblasts and *in vitro* differentiated myotubes. DNA-FISH, which includes RNase treatment and denaturation of cellular DNA, yielded two different signal patterns that were dependent on the expression status of the studied genes. In expressing cells, dispersed signals consisting of several small foci of various sizes were observed in the nuclear interior, suggesting that transcribed genes are strongly decondensed (Fig.1C). In non-expressing cells, e.g. fibroblasts, genes were condensed into single considerably larger and brighter foci sequestered to the nuclear periphery (Fig.1B, S2), in agreement with their positioning within or close to LADs identified in mouse fibroblasts (Peric-Hupkes et al., 2010).

Next, we performed RNA-FISH on the same set of samples, omitting both RNase treatment and cell DNA denaturation. With this setup, the same genomic probes as used for DNA-FISH hybridized exclusively to nascent RNA (nRNA) transcripts. As expected, in cells not expressing the genes, FISH signals were absent. In striking difference, expressing cells exhibited massive RNA signals (Fig.1D, S2). Remarkably, the RNA-FISH signals were extended and had shapes of coiled loops, especially prominent in case of the *Tg* gene where they were spread throughout the nuclear interior and measured up to 10 *µm* (Fig.1D, S2A, SI Movie 1).

Almost each respective nucleus exhibited two RNA-FISH signals corresponding to two gene alleles. Signals varied in shape and coiling degree and no consistent pattern of loop folding or location was observed with exception to invariable interior positioning. Since introns, as a rule, are substantially longer than exons, we assume that the used genomic probes were hybridizing mostly to unspliced introns of nRNAs attached to elongating RNAPIIs, thereby outlining the contours of expressed genes. Structures revealed by RNA-FISH resembled the lateral loops of lampbrush chromosomes with massive ongoing transcription and we further refer to these structures as Transcription Loops (TLs).

Importantly, a combination of both substantial length and high expression of a gene is essential for TL visualization. Long but weakly expressed genes, as well as highly expressed but short genes, do not form microscopically resolvable loops, as we tested by FISH for ca. 50 mouse genes in various cell types of tissues and cultured cells (see examples in SI Table S4).

Transcription Loops manifest the progression of transcription and co-transcriptional splicing

To confirm that the observed RNA signals are not accumulations of messenger RNAs but represent nRNA transcripts, we performed two RNA-FISH experiments. First, we used differentially labeled probes hybridizing to 5’ exons (exons 2-12) and 3’ exons (exons 33-47) of the *Tg* gene (Fig.2A1). We reasoned that if loops represent mere mRNA accumulation, both probes would label them, however in case of ongoing transcription, the probe for 3’ exons would label only the second half of TLs (Fig.2A2,3). RNA-FISH with differentially labeled probes for 5’ and 3’ exons labeled *Tg* TLs according to the second scenario (Fig.2A4), confirming that TLs are decorated with growing nRNAs.

Secondly, we used genomic probes highlighting 5‘ and 3‘ introns of the *Tg* gene (Fig.2B1). We expected that the two probes hybridize to nRNAs on TLs in a consecutive manner as a result of co-transcriptional splicing (Fig.2B2,3) and, indeed, observed such a pattern (Fig.2B4): the 5’ probe strongly labeled the first half of the loop by hybridizing to nRNAs containing yet unspliced 5’ introns and the 3’ probe labeled the second half of the loop by hybridizing to yet unspliced 3' introns. Co-transcriptional splicing was shown for all other studied TLs using RNA-FISH with respective genomic probes highlighting introns in the 5' and 3' halves of the respected genes (Fig.2C). Although previous research has already provided evidence for co-transcriptional splicing employing microscopy (Carmo-Fonseca and Kirchhausen, 2014), the use of long highly expressed genes in this study allowed for the first time to unambiguously resolve this process in space.

In addition to TLs, RNA-FISH revealed numerous RNA granules scattered throughout the entire nucleoplasm of the studied nuclei (Fig.S3A). These granules may represent either accumulations of excised but not yet degraded introns or mature mRNAs. For the *Ttn* gene, whose mRNAs are exceptionally long (ca. 101 kb), the second interpretation is more probable, especially because similar granules were also present in the myotube cytoplasm (Fig.S3A).

Next, we asked what the degree of chromatin compaction of highly expressed genes, such as *Tg*, is. Research by others suggests that highly expressed sequences, such as ribosomal genes in amplified nucleoli of amphibian oocytes (Miller and Beatty, 1969) or sequences on lateral loops of lampbrush chromosomes (Kaufmann et al., 2012), are depleted of nucleosomes as a result of a high polymerase occupancy. Nucleosome eviction is suggested for highly expressed genes (Kulaeva et al., 2010), whereas a moderate expression is apparently compatible with nucleosome retention (Farnung et al., 2018; Gaykalova et al., 2015; Liu et al., 2020). To estimate the compaction level of *Tg* TLs we measured the contour length of three *Tg* gene regions after RNA-FISH and found that the *Tg* compaction level is ca. 17 kb/µm and thus compatible with nucleosomal chromatin (Fig.S3B). However, since these measurements were performed on 2D projections of confocal stacks and since *Tg* TLs are strongly coiled (Fig.S3C,D), the measured compaction of the *Tg* gene was rather overestimated and we cannot exclude that a significant proportion of nucleosomes is lost from *Tg* TLs.

Collectively, the above experiments show that TLs are formed by highly decondensed genes with ongoing transcription and co-transcriptional splicing.

Transcription Loops are open loops with separated flanks

It is widely believed that physical interactions between the TSS and TTS are crucial for a high level of expression. Such interactions have been interpreted as a 'bridge' for RNAPIIs enabling them to immediately reinitiate transcription after termination, i.e. to "transcribe in circles" (Ansari and Hampsey, 2005). To assess whether the beginnings and the ends of the selected long highly expressed genes reside in close proximity to each other, we visualized TLs and their genomic flanks for *Tg*, *Ttn* and *Myh11* genes and found that the 5' and 3' gene flanks are microscopically separated in 85%, 87% and 90% of the respective alleles. In particular, 3D distance measurements between flanks showed that they can be separated by up to 2.2 *µm* (*Tg*) and 2.9 *µm* (*Ttn*), and that the inter-flank median distances in expressing cells are ca. 2-fold larger compared to control cells not expressing these genes (Fig.2D). This finding demonstrates that contrary to the "transcription cycle" hypothesis, TLs of long highly expressed genes are open loops with separated flanks.

Transcription Loops of long genes are excluded from harboring chromosome territories and can deform them

Chromosome territoriality is a central doctrine of nuclear organization (Cremer & Cremer 2010). To investigate the relationship between TLs and their harboring chromosomes, we visualized the *Tg* TL and chromosome 15 by FISH detecting both DNA and RNA. We found that the *Tg* TL is excluded from the chromosome territory and either expands into the nucleoplasm or coils next to the chromosome, forming its own “transcription territory” (Fig.3A). The density of nRNAs and the associated protein machinery within such a “transcription territory” is so high that chromatin is largely excluded from the nuclear regions occupied by the TLs (Fig.3B). The highly upregulated *Tg* gene exhibited territorial exclusion in 100% of alleles; less expressed *Ttn* and *Neb* were excluded from chromosome 2 in 90% and 72% of alleles, respectively, confirming previous observations that expressed genes occupy peripheral positions within chromosomes or even extend away from them (Cremer & Cremer 2010).

Remarkably, *Tg* TLs split chromosome 15 territories into two halves in 2% of *Tg* alleles. The clear gaps in the painting signal were marked with the 5’ and 3’ *Tg* flanking sequences and filled with *Tg* TLs (Fig.3C). Chromosome 15 is noticeably asymmetrical in respect of its transcriptional activity (Fig.S4A1,B), and therefore is radially stretched in thyrocyte nuclei (Fig.S4A2,3). The high expression of the *Tg* gene, positioned at the border between the active and inactive regions (Fig.S4A1), presumably facilitates stretching further, which results in chromosome splitting (Fig.3C).

Taken together, these microscopy data demonstrate that transcription of long genes can generate considerable mechanical forces sufficient to relocalize these genes, break and reorganize chromosomal territories.

Hi-C analysis reveals local chromosome reorganization

Chromosome paints are low-complexity probes and therefore we employed Hi-C to further study the arrangement of TLs within their harboring chromosomes. We reasoned that if an expressed gene loops out of its chromosome, its *cis*-contacts are diminished compared to the *cis*-contacts of the same gene in a silent state, i.e. the *cis*-to-*trans* contact frequency ratios (further referred to as *cis/trans* ratios) must be low. To this end, we performed Hi-C analysis of mouse thyroid, bladder, cultured myoblasts and myotubes differentiated from them. First of all, we observed that *cis/trans* ratios are generally lower in the A than in the B compartment (Fig.S4B,C), which is in agreement with a previously reported positive correlation between interchromosomal contact probability and transcriptional activity (Kalhor et al., 2012), and can be tentatively attributed to enrichment of the A compartment in expressed genes looping out from chromosomes.

Analysis of the *cis/trans* ratios revealed that genes, exhibiting the largest TLs, such as *Tg*, *Ttn* and *Myh11*, display a dip in the *cis/trans* ratio profile, which indicates that they preferentially contact other chromosomes, while the same genes in a silent state and other long but weakly expressed genes don’t exhibit such a dip (Fig.3D). The *Neb* and *Cald1* genes, characterized by lower expression and visibly smaller TLs, have *cis/trans* ratios similar to those of other long lowly expressed genes (Fig.3D). In agreement with the observation that highly expressed long genes loop out of chromosomes, we found that none of the selected genes suppresses contacts across them, either in the “*off*”, or in the “*on”* states (Fig.S5). Thus we concluded that the effect of TL formation is on average localized to the immediate vicinity of a gene and leaves most of the harboring chromosome largely unaffected.

In summary, both microscopy and Hi-C analysis confirm transcription-dependent looping of highly expressed long genes out of their harboring chromosomes.

Transcription Loops are dynamic structures

To confirm that TLs are dynamic structures formed upon transcriptional activation, we performed several experiments to activate or inhibit *Ttn* transcription *in vitro*. First, we aimed to induce *Ttn* TLs in cells that do not express this gene and generated myoblasts stably expressing dCas9 conjugated with the tripartite transcription activator VP64-p65-Rta (VPR) (Chavez et al., 2015). After transfection of these cells with a cocktail of six gRNAs targeting the *Ttn* promoter region, expression of *Ttn* in myoblasts was induced to a level comparable to myotubes (Fig.4A2) and *Ttn* TLs emerged in about 90% of transfected cells (Fig.4A1). TL formation was accompanied by a 2-fold increase of *Ttn* median inter-flank distances (Fig.4A3).

Conversely, we aimed to eliminate *Ttn* TLs in differentiated myotubes by transcription inhibition. To this end, we incubated myotubes with α-amanitin or actinomycin D and demonstrated that both drugs cause the disappearance of TLs (Fig.4B1). In agreement with different mechanisms of their action (Bensaude, 2011), the drugs differently affected gene body condensation. In case of α-amanitin, which prevents DNA and RNA translocation by binding near the catalytic site of RNAPII, the *Ttn* gene condensed and the gene flanks converged. In case of actinomycin D, which prevents RNAPII progression by intercalating into the DNA minor groove, *Ttn* remained decondensed with separated flanks (Fig.4B2-4).

To monitor TL dynamics further, we treated differentiated myotubes with 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), a drug that reversibly inhibits CDK9 and thereby keeps RNAPII in a stalled state close to the TSS, preventing transcription elongation (Bensaude, 2011). As anticipated, during the first 2 hours of drug treatment, we observed gradual gene body condensation and reduction of the RNA-FISH signal sizes of nRNAs that were attached to still elongating RNAPIIs (Fig.4C1). After drug removal, we observed restoration and growth of the loop signal, in particular, convincingly demonstrated by differential labeling of 5' and 3' nascent RNAs (Fig.4C2) with the 5' signal emerging first and the 3' signal emerging only after ca. 1 h. Treatment with DRB dynamically modified *Ttn* inter-flank distances: flanks converged upon drug addition and diverged after drug removal (Fig.4C3).

Collectively, the above experiments demonstrate that TLs are dynamic structures, which are formed by genes upon transcription activation but withdrawn and condensed upon silencing.

Highly expressed genes are expressed biallelically and exhibit transcriptional bursting

Compelling evidence has been provided for random monoallelic expression of autosomal not imprinted genes in various cell types (Savova et al., 2016; Savova et al., 2013). Scoring of RNA-FISH signals revealed that all five selected genes are expressed mostly biallelically, which is in line with the observation that monoallelic expression is more common for lowly expressed genes (Deng et al., 2014). For example, only 9% of cultured myoblasts and 5.4% of myotubes had a single *Cald1* or *Ttn* RNA signal, respectively. In tissues, the proportions of nuclei with a single active allele were even lower: 2% for *Ttn* in skeletal muscle, 3% for *Tg* in thyroid and 5% for *Myh11* in colon (Fig.S6A-D).

Transcriptional bursting of eukaryotic genes is the universal principle of gene expression with burst frequency ranging from minutes to hours and burst size from one to hundreds of transcripts (Chubb et al., 2006; Li et al., 2019; Nicolas et al., 2017). Given an RNAPII elongation speed of 3.8 kb/min (Singh and Padgett, 2009), short genes (<20 kb) are expressed within seconds to minutes and, therefore, at any given moment, a cell with a highly expressed short gene can have either both alleles active or pausing, or only one allele active. In case of long genes, transcription takes longer (e.g., 75 min for *Ttn* and 50 min for *Tg*) and one can expect multiple transcription bursts along the gene axis. In agreement with this and as discussed above, both alleles of the selected long genes were detected by RNA-FISH in the majority of the cells.

Expression level is directly related to the frequency and duration of transcription bursts (Larsson et al., 2019). Since the expression levels of the studied long genes are different, we expected that they exhibit different bursting patterns and that these patterns are microscopically detectable. First of all, the sizes of TLs are disproportionate to the gene length: e.g., *Ttn* (279 kb) RNA-FISH signals are less expanded compared to those of the *Tg* gene (180 kb) (Fig.1D; 5A,B). Secondly, the studied genes displayed different patterns of gene condensation: e.g., the gene body of *Tg* (22,924 TPM) was hardly traceable and marked by few small faintly labeled foci, whereas the 10-fold less expressed *Ttn* gene (2,110 TPM) displayed a series of several larger foci with small gaps in between (Fig.1C; 5A,B). The *Neb* gene, which is expressed more than twice as low as *Ttn*, typically exhibited higher gene body condensation with only 1 or 2 microscopically resolvable gaps, suggesting that this gene is transcribed in even shorter bursts separated by longer pauses (Fig.1C; 5C). The body of *Dmd*, the longest gene (ca. 2.3 Mb) with a very low expression (5 TPM), remained condensed and folded during transcription (Fig.5D).

We presume that the granules and gaps detected by DNA-FISH correspond to transcription pauses and bursts, respectively (Fig.5; SI Movie 2). Accordingly, the highly expressed *Tg* is transcribed in long bursts interrupted by short pauses, whereas the more lowly expressed *Ttn* is transcribed in short bursts separated by long pauses (Fig.5A,B). To test the differences in bursting of the studied genes, we detected nRNA transcripts over 30 kb of the *Tg*, *Ttn* and *Cald1 genes* and scored alleles with and without the signals, arguing that the absence of a 30 kb RNA-FISH signal in a particular allele indicates that this region falls into a transcription pause. We found that 2.3%, 11.6% and 14.4% of *Tg*, *Ttn* and *Cald1* alleles, respectively, were indeed missing the 30 kb signal indicating that a transcription pause can last over 30 kb and that *Tg* predictably exhibits several folds fewer such alleles than *Ttn* or *Cald1* (Fig.S6E).

In summary, we conclude that the level of gene expression is manifested in a microscopically detectable pattern of alternating transcription bursts and pauses.

Transcription Loops are stiff structures

During a transcription burst, RNAPIIs traveling along a DNA-template are tightly spaced into a convoy (Fujita et al., 2016; Tantale et al., 2016). As discussed in the previous paragraph, the higher the expression, the longer the transcription bursts and the longer the RNAPII convoys. Every RNAPII complex is associated with a nascent ribonucleoprotein complex (nRNP) consisting of synthesized nRNA and numerous proteins involved in RNA processing, quality control, transport, translation, etc. (Muller-McNicoll and Neugebauer, 2013) (Fig.6A). Therefore, nRNPs are voluminous structures, which in case of long nRNAs, can exceed the size of nucleosomes (10-11 nm; (Olins and Olins, 1974)) and even RNAPIIs (13-14 nm; (Liu et al., 2018)). These considerations lead us to the hypothesis that dense decoration with voluminous nRNPs changes the physical properties of a highly expressed gene converting it into a stiffer structure compared to a silent gene. The stiffness, in turn, forces a gene to separate its flanks and protrude from a harboring chromosome into the nuclear space, in other words, to form a TL (Fig.6A; SI Movie 2). Below we provide supporting evidence to the hypothesis, including computer simulations (see next section).

Firstly, nRNP granules formed on long genes are indeed voluminous structures. The widely acknowledged example of unambiguously identified nRNPs formed over known genes with a known length, are BR genes (BR1, BR2.1, BR2.2, BR6), corresponding to Balbiani rings on polytene chromosomes of *Chironomus tentans*. BR genes are 35-40 kb long and include only four small (≤ 1kb) introns. The RNP granules formed on the genes increase in size from 5' to 3' end with diameters increasing from 20 nm to 50 nm, respectively (Bjork and Wieslander, 2015; Olins et al., 1983). Such voluminous nRNPs densely arranged along transcribed BR genes can convert them into stiff structures, which indeed "puff" away from polytene chromosomes in a manner similar to looping out TLs described in this study.

Secondly, formation of large nRNPs undoubtedly requires long genes, since short genes produce short nRNAs, corresponding to minute nRNPs, which do not prevent gene flexibility and coiling (Fig.6B,C3). Large nRNPs can form in two ways: either on genes with long exons decorated by nRNAs growing from the 5' to 3' gene end (Fig.6C1), or on genes with long introns with nRNAs growing from 5' to 3' splice sites (Fig.6C2). In both cases, substantial gene regions are decorated by long nRNAs corresponding to voluminous nRNPs imposing stiffness and thus gene expansion. The *Ttn* gene, from which a 101 kb mRNA is read, can serve as an example of the first case (Fig.6B): the *cis/trans* ratio curve for *Ttn* displayed an asymmetry with a drop towards the 3' end (Fig.3D), indicating a stronger exclusion from the harboring locus and less coiling of the gene's 3' end. The *Tg* gene includes a 54-kb intron (*intron 41*), over which mRNAs grow from ca. 6 to 60 kb, and thus might represent the second case. Indeed, RNA-FISH with oligoprobes selectively labeling this intron reveals a gradient of the signal thickness, which reflects the growth of nRNPs from 5' to 3' intron splice sites (Fig.S3D). Moreover, simultaneous labeling of intron 41 and other gene regions detected a disproportionately larger extension of *Tg* TLs over the intron region in comparison to the rest of the gene (Fig.6D).

Thirdly, we reasoned that splicing inhibition must lead to lengthening of nRNAs and hence to increasing of nRNPs' sizes, eventually facilitating TL expansion. To this end, we inhibited splicing with Pladienolide B (Kotake et al., 2007) in cultured myoblasts expressing *Cald1*, which comprises three long introns at the 5' end, and then performed RNA-FISH with genomic probes for the intron-rich 5' end and the exon-rich 3' end of the gene (Fig.6E). Despite the massive abortion of transcription upon Pladienolide B treatment, manifested by the disappearance of the 3' FISH signal and multiple nucleoplasmic accumulations of nRNAs, we still observed an expected increase of the 5' part of *Cald1* TLs size by 2.5-fold (Fig.6E).

Polymer modeling confirms stiffness of TLs

Assessing gene stiffness is a difficult or maybe even impossible biophysical experiment. Therefore, next we turned to polymer models to investigate whether gene stiffness indeed can give rise to formation of TL, with their enormous sizes, large separation of the flanks, expansion beyond the chromosome territories, and observed Hi-C features.

We modeled chromosomes as 50 Mb polymers with a monomer size of 1 kb, roughly corresponding to 5 nucleosomes arranged in a 20 nm globule. We generated 6 territorial chromosomes confined to a spherical nucleus by initiating them from a mitotic-like conformation and letting them expand (Fig.7A1; for other model parameters see Fig.S7A). On each chromosome, we assigned a 100 kb region as the gene of interest (Fig.7A2), and explored at which simulation parameters we can recapitulate biological observables for the *Tg* gene, including visual appearance of TLs, distances between flanks and *cis/trans* contact ratio along the gene length (Fig.7B) obtained from simulated Hi-C (Fig.7C-F, S7C).

First, we tried to reproduce TL formation by increasing the gene contour length, reasoning that high expression leads to lengthening *via* nucleosome eviction. We assumed that complete loss of nucleosomes leads to a ≈15-fold increase of the gene contour. However, simulated genes at this condition appear as rather amorphous structures without a discernible contour and thus do not reproduce the visual appearance of TLs. Although median inter-flank distances of simulated genes were comparable to distances measured in thyrocytes (0.88 µm versus 0.75 µm, respectively) the *cis/trans* ratio of elongated genes was not different from the rest of the simulated chromosomes (Fig.7D). Therefore, we concluded that a mere increase of the gene contour length due to nucleosome loss cannot accurately recapitulate the formation of a TL.

Next, we reasoned that due to dense decoration with nRNPs (see above), a highly transcribed gene has an increased bending stiffness compared to the rest of the chromosome. We increased the contour length by 3-fold to account for partial nucleosome loss and stretching induced by the decoration with nRNPs. Then we altered the stiffness of our simulated genes and found that a 12-fold increase in stiffness reproduces the experimental features of *Tg* TLs, including visual appearance, inter-flank distances, and low *cis/trans* ratio of Hi-C contacts (Fig.7E). Moreover, the model successfully recapitulates the dynamic behavior of TL flanks upon gene silencing (Fig.S7B). In an effort to account for the gradual increase of nRNPs along the gene, we introduced a tapered stiffness profile along a gene with stiffness gradually increasing from 1 to 24-fold (Fig.S7C). The added gradient slightly reduced the TL's inter-flank distance, lead to a visibly more curled 5' end and a more stretched 3' end of the simulated TLs (Fig.7F), and thus recapitulated the morphology of *Tg* TLs observed microscopically. Furthermore, the stiffness gradient led to asymmetry in the *cis/trans* profile along the gene, in remarkable agreement with experimental Hi-C data (Fig.3D,7F).

We point out that in all simulated conditions the dramatic change in visual appearance of the TLs is not accompanied by significant changes in the Hi-C maps with exception to short-range effects (Fig.S7C). In particular, the formation of TLs does not lead to significant insulation (Fig.S5). These findings, though surprising, agree with experimental data. Finally, we point out that our simulations serve as a proof of principle that increased stiffness can explain TL formation. We do not claim that parameters like the fold-increase in stiffness and contour length can be determined accurately from our current simulations.

In conclusion, polymer modeling of highly transcribed long genes corroborates our hypothesis about TL stiffness and explains the phenomenon of TL exclusion from harboring chromosomes and their expansion into the nuclear space.

DISCUSSION

Investigations about the spatial arrangement of transcribed genes have been limited by two factors. On the one hand, most of the previously studied highly expressed genes are short (e.g., (Schoenfelder et al., 2010) and thus cannot be resolved by light microscopy. On the other hand, the studied long genes (e.g., (Schoenfelder et al., 2010) are not sufficiently highly expressed to have their axes densely decorated with nRNAs, which is a prerequisite for the visualization of decondensed genes. In this work, we used long and highly expressed genes as a magnifying glass allowing us to elucidate the structure of transcribed genes.

First, we show that an expressed gene forms a TL with RNAPIIs moving along the gene axis and carrying nRNA transcripts that undergo co-transcriptional splicing (SI Movie 2). The observed sequential patterns of exon and intron labeling along TLs (Fig.2A,B) rule out the possibility of gene expression in a transcription factory with immobilized RNAPIIs and nRNA concentrated in one spot (Cook, 1999). Moreover, the geometry of highly extended TLs is also not compatible with a stable contact between an enhancer-promoter complex and a gene body during transcription (Lee et al., 2015; Zheng et al., 2019). Noteworthy, although TLs are visually similar to lateral loops of lampbrush chromosomes, they do not exhibit a characteristic gradient of nRNPs along their axes (Macgregor, 1993) (Fig.1D), evidently, due to co-transcriptional splicing occurring along TLs and loops coiling. Such a gradient, however, is readily manifested by expanded 54 kb *Tg* introns, over which nRNAs grow steadily from ca. 6 to 60 kb before the intron is spliced out (Fig.S3D).

Secondly, we demonstrate that TL formation dynamically modifies the harboring chromosomal locus: transcription activation causes gene flank divergence and extension of the gene body from the chromosome into the nuclear space, whereas transcription silencing causes gene flank convergence and condensation of the gene body (Fig.3,4; SI Movie 2). The fact that TLs are open loops with separated gene flanks in the majority of alleles argues against the proclaimed necessity of TSS - TTS association for maintenance of a high level of transcription (Hampsey et al., 2011; Martin et al., 2005; Singh and Hampsey, 2007). The spreading of TLs over euchromatic nuclear regions by extending away from their harboring chromosomes and even breaking them apart, challenges the importance of chromosome territoriality in transcription regulation (Cremer and Cremer, 2010). A greater number of trans-contacts, as seen in our Hi-C analysis, for both euchromatin in general (Fig.S4C) and long highly expressed genes in particular (Fig.3D), indicates that expressed genes can extensively interact with the surrounding euchromatin. These data, together with previous descriptions of intermingling interphase chromosomes (Branco and Pombo, 2006), indicate that chromosome territoriality is not so much an intrinsic functional feature of interphase nucleus as it is a mere consequence of the last mitosis (Solovei et al., 2016), an event for which genome partitioning into separate chromosomes is utterly crucial.

Thirdly, we demonstrate that TLs consist of alternating decondensed and condensed gene regions and speculate that this pattern corresponds to transcription bursts and pauses, respectively (Fig.5; SI Movie 2). The discontinuous process of transcription with pulses and pauses has been accepted as a basic rule of gene expression in both pro- and eukaryotes (Nicolas et al., 2017) and has been considered a factor increasing the sensitivity in the control of gene expression (Chubb et al., 2006). Although direct measurements of bursting were not possible in our work, based mostly on fixed tissues and cells, matching of expression levels and observed patterns of condensed and decondensed gene regions suggests that high expression corresponds to long bursts interrupted by short pauses (e.g., *Tg*) and, alternatively, lower expression corresponds to shorter bursts and longer pauses (e.g., *Ttn*). This notion is supported by a higher proportion of *Ttn* and *Cald1* alleles lacking RNA-FISH signal over a 30 kb region in comparison to the more highly expressed *Tg* gene (Fig.S6E).

And finally, to explain the displacement of TLs away from their harboring loci, we put forward a hypothesis of increased stiffness of TLs, which is caused by a dense decoration of the gene axis with nRNPs progressively growing in size (Fig.6A; SI Movie 2). First of all, the hypothesis is supported by electron microscopy data on size of RNPs formed on long exon-rich genes, such as BR genes forming Balbiani rings in polytene chromosomes (Bjork and Wieslander, 2015; Olins et al., 1983). The stiffness hypothesis is also successfully tested in this work: Hi-C analysis exhibits asymmetrical exclusion of the exon-rich *Ttn* 3' end from the chromosomal context (Fig.3D); FISH data show disproportional expansion of *Tg* TL region corresponding to the 54 kb long intron (Fig.6D); the intron-rich portion of the *Cald1* TL increasingly expands upon splicing inhibition (Fig.6E). This hypothesis also explains why highly expressed but short genes do not form resolvable TLs: a short gene lacks long introns and is thus decorated by small nRNPs permitting gene-axis flexibility compatible with gene coiling (Fig.6B).

Importantly, the stiffness hypothesis is successfully tested by modeling a transcribed gene as a region of increased bending rigidity within a bigger chromosome. Such simulations recapitulated the features of *Tg* TLs (Fig.7), namely (i) a strong increase in size with a clearly discernible contour, (ii) an increase in gene flank distance, and (iii) a preference to contact other chromosomes leading to a dip in the *cis/trans* profile within the gene body. Our simulations of TLs as stiff regions also reproduce the dynamic swelling of a gene upon activation (increasing stiffness) and recondensation upon deactivation (decreasing stiffness), as well as the surprising absence of long-range insulation across TLs as observed in Hi-C. In summary, we found that the relatively simple mechanistic model qualitatively explains a number of different features of TLs and thus strongly supports the hypothesis of increased stiffness as the mechanism for TL formation. More realistic scenarios of models including nRNPs, travelling RNAPIIs and co-transcriptional splicing, may allow us to further clarify the mechanics of TL formation in the future.

Collectively, our data suggest that TL formation is a universal principle of eukaryotic gene expression (SI Movie 2) that has not been appreciated until now due to the limited resolution of light microscopy in combination with small size or low expression of the studied genes.

SUPPLEMENTAL DATA

Supplemental data include STAR Methods, 7 Supplemental figures, 2 SI movies and 5 SI Tables

ACKNOWLEDGEMENTS

We are grateful to David Hörl and Joel Ryan for help with ImageJ plugins and programming. We thank Christopher Mulholland for RNA-seq library preparation. We acknowledge Jack Bates for fruitful discussions, and Dimitar Kralev and Tamako Suzuki for help with animation preparation. This work has been supported by the Deutsche Forschungsgemeinschaft grants (SO1054/1 and SP2202/SO1054/2 to IS; SFB1064 to IS and HL, SPP 2202/LE721/17-1 to HL) and by National Institutes of Health grants (HG007743 to HL; HG003143 to JD; GM114190 to LM by the Center for 3D Structure and Physics of the Genome of NIH 4DN Consortium, DK107980). JD is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

IS conceived the project. SLe, YF, KT, IS obtained biological samples. IS, YF, SLe, SU conceived and performed microscopic experiments and image analysis. SLe and SLi performed ChIP-seq experiments; SB performed RNA-seq and ChIP-seq analyses. YF and EH performed Hi-C experiments; EH, JD, JN and LM performed Hi-C analysis. JN with contribution from LM performed simulations. SLe, SU, JN and IS prepared the figures. IS wrote the manuscript with contributions from SLe, YF, JN, EH, HL, LM and JD.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Table S1. Excel spreadsheet includes distribution of human genes according to their length, the list of protein-coding genes and their expression levels in different human tissues, and the list of selected genes with length longer than 100 kb and expression level of ca. 1,000 TPM.

Table S2. Excel spreadsheet includes distribution of mouse genes according to their length, the list of protein-coding genes and their expression levels in two mouse tissues and two cultured cell types, list of five selected genes with length of ca. 100 kb and expression level of ca. 1,000 TPM and list of long lowly expressed genes used for comparison of *cis/trans* contact ratios with selected genes

Table S3. Word table with the list and schematics of BAC clones used for DNA- and RNA-FISH experiments.

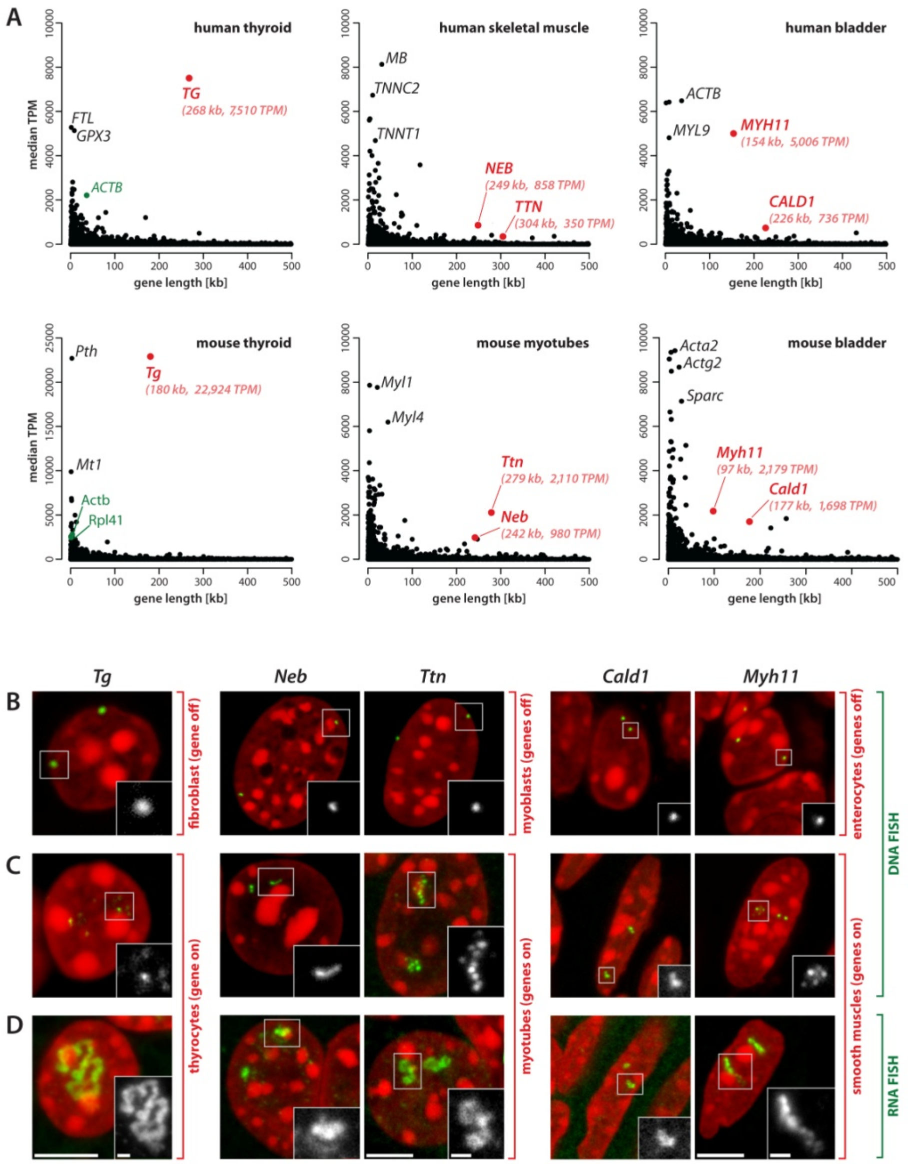
Table S4. Word file with the list of long lowly expressed (S4A) and short highly expressed genes (S4B) not forming resolvable TLs, supplemented with corresponding FISH images.

Table S5. Excel spreadsheet containing primer pairs used for (1) introduction of protospacer sequences into U6-gRNA-GFP-H2A; (2) generation of FISH probe for 30 kb at the *Tg* 5' end; (3) generation of FISH probe for *Tg* intron 41; (4) generation of FISH probe for 30 kb at the *Ttn* 3' end.

Movie 1. Confocal stacks through nuclei of mouse thyrocytes (counterstained with DAPI, *red*) after RNA-FISH with genomic probe for the *Tg* gene (*green*). Note the volatile shape of the *Tg* TLs and their great expansion into the nuclear interior.

Movie 2. Cartoon showing how transcription initiation and termination of a highly expressed gene lead to formation and disappearance of a TL.

The sequence of events: a gene body (*orange thread*) is coiled within a locus (*blue threads*) in a compact structure; upon transcription initiation, RNAPIIs (*dark grey oval structures*) are loading at the gene promoter (*red*) and begin to elongate; during elongation, nRNPs appear and grow in size (*depicted as grey amorphous structures*); during a transcription pause, chromatin of the gene is coiled and forms a sliding knot (*orange thread*), dynamically formed beyond the last RNAPII of the first burst and disentangled by first RNAPII of the second burst; dense decoration with voluminous nRNP rigidifies gene axis and forces its expansion, as well as divergence of gene flanks (*blue threads*); RNAPIIs of the first transcription burst reach the 3' gene end, release attached nRNPs and dissociate from the gene; the sliding chromatin knot reaches the 3' gene end; RNAPIIs stop loading at the promoter at the end of the second burst; gene condensation starts 5'-terminally; the gene flanks begin to converge; the gene returns to its coiled compact state.



FIGURES

Figure 1. Highly expressed long genes form transcription loops (TLs).

A, Analysis of gene expression in selected human tissues (*top*) retrieved from the GTEx database and RNA-seq analysis of corresponding mouse tissues and cells from this study (*bottom*). Expression level (median TPM) is plotted against gene length according to GENCODE. Candidate genes with a length of ca. 100 kb or longer and an expression level of ca. 1,000 TPM or above are marked in *red*. Note the exceptionally high level of *Tg* expression, exceeding expression of housekeeping genes, such as *Actb* and *Rpl41* marked in *green*.

B-D, Five selected genes after either DNA- or RNA-FISH. Control cells not expressing the respective genes exhibit focus-like condensed DNA signals at the nuclear periphery (B). As a result of high gene decondensation during transcription, DNA-FISH fails to outline gene bodies and reveals only several condensed foci (C). RNA-FISH highlights gene contours (TLs) due to hybridization to many nRNAs decorating the genes (D).

Insertions show enlarged gray scale images of respective TLs. Images are *projections of 1-2.5 µm* confocal stacks. Scale bars, *5 µm*; in insertions, *1 µm*.

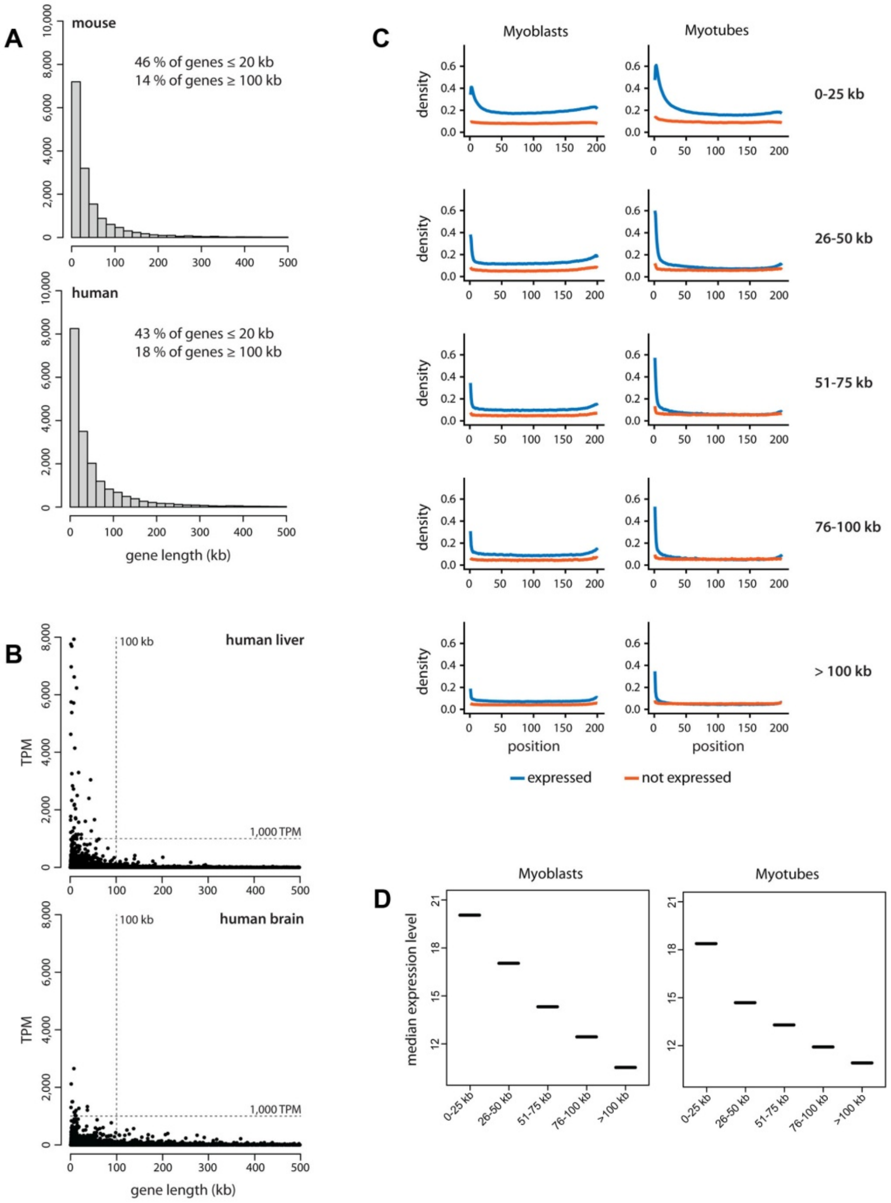


Figure S1 (related to Figure 1). Long genes are rare and expressed at a lower levels than short genes.

A, Analysis of gene length distribution within the human and mouse genomes showed that about 43% and 46% of all protein coding genes, respectively, have a length below 20 kb and only 18% and 14% have a length of 100 kb or above. Histograms of the gene length distribution in mouse (*top*) and human (*bottom*) with bin size of 20 kb. Genes are annotated according to GENCODE. Only genes with a length below 500 kb are shown on the graphs.

B, To select suitable genes for visualization we studied gene expression profiles across 50 human tissues using the publicly available Genotype-Tissue Expression database (GTEx Consortium; https://www.gtexportal.org/home/) and found that long genes are generally not highly expressed. For example, in the brain (*bottom*) and liver (*top*), there were no genes with both a length of 100 kb or longer and with an expression of 1,000 TPM or above. Genes were annotated according to GENCODE. Expression level (median TPM) is plotted against gene length (kb).

C, Expressed short genes exhibit a higher RNAPII occupancy than expressed long genes. ChIP-seq with an antibody against the CTD of RNAPII in cultured mouse myoblasts (Pmi28) and *in vitro* differentiated myotubes. All genes, expressed (>1 TPM, *blue*) and silent (<1 TPM, *red*), were split into five categories according to their size. RNAPII density (*Y-axis*) is plotted against gene length (*X-axis*); each gene is divided into 200 equally sized bins and genes from the same size category are aligned according to the bins. Expressed genes display a higher occupancy with RNAPII compared to non-expressed genes, especially in the TSS region. In the group of expressed genes, the RNAPII occupancy negatively correlates with gene length: the shorter the genes, the higher the RNAPII occupancy.

D, Analysis of RNA-seq data for myoblasts (left) and myotubes (right). The median expression level is higher in groups containing shorter genes (<25 kb) and generally negatively correlates with gene length.

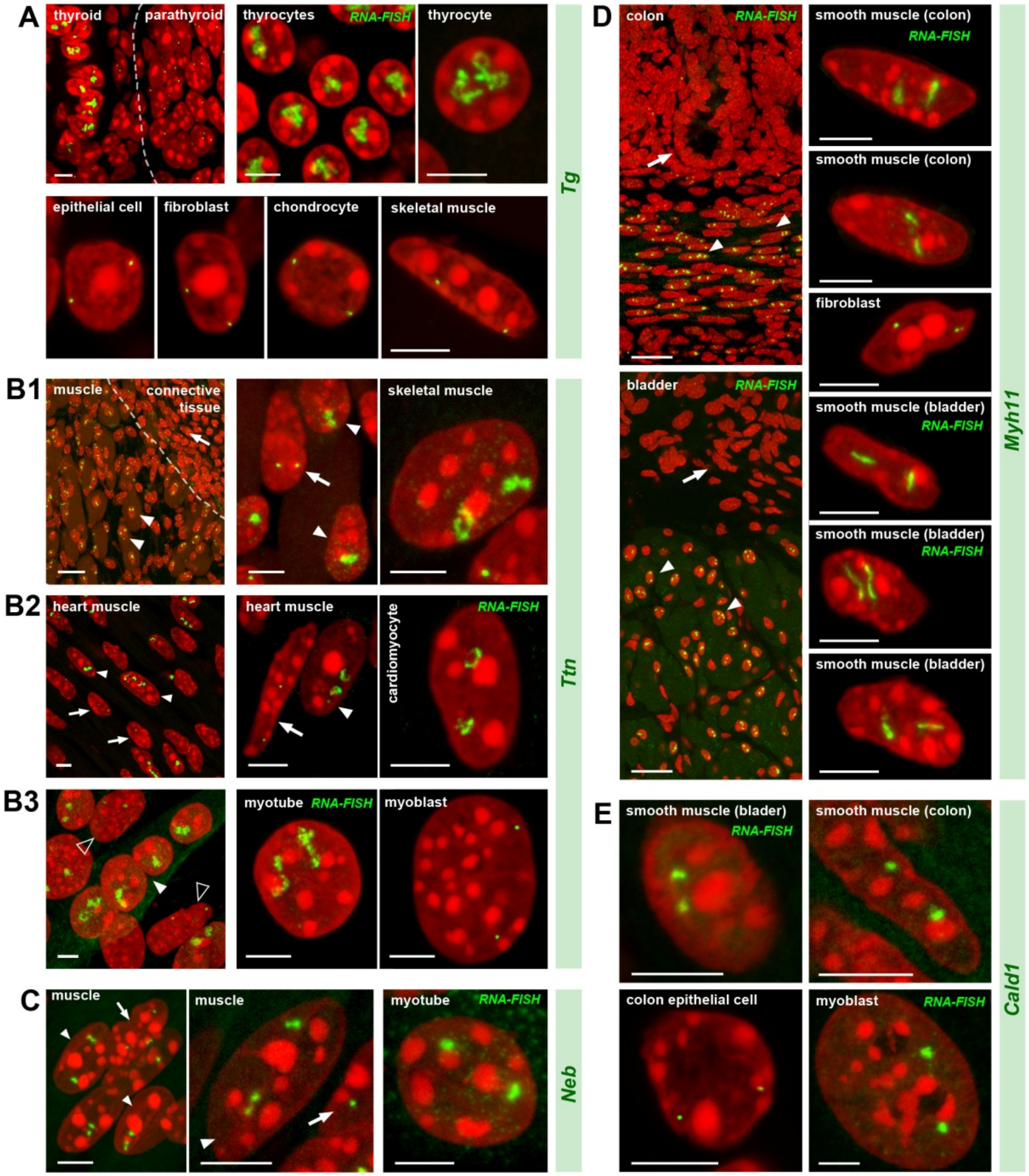

A
thyroid
parathyroid
thyrocytes
RNA-FISH
thyrocyte
epithelial cell
fibroblast
chondrocyte
skeletal muscle
Tg
B1
muscle
connective tissue
skeletal muscle
B2
heart muscle
heart muscle
cardiomyocyte
RNA-FISH
B3
myotube
RNA-FISH
myoblast
Ttn
C
muscle
muscle
myotube
RNA-FISH
Neb
D
colon
RNA-FISH
smooth muscle (colon)
RNA-FISH
smooth muscle (colon)
fibroblast
bladder
RNA-FISH
smooth muscle (bladder)
RNA-FISH
smooth muscle (bladder)
RNA-FISH
smooth muscle (bladder)
Myh11
E
smooth muscle (blader)
RNA-FISH
smooth muscle (colon)
colon epithelial cell
myoblast
RNA-FISH
Cald1

Figure S2 (related to Figure 1). Visualization of the five selected genes in expressing and not expressing cells.

A, The *Tg* gene is expressed in thyrocytes where both alleles form prominent TLs expanded into the nuclear interior. In neighboring cells with a silent *Tg* gene - parathyroid gland cells, tracheal chondrocytes, epithelial cells, fibroblasts and muscles - *Tg* is highly condensed and sequestered to the nuclear periphery.

B, The *Ttn* gene is expressed in skeletal muscle (B1), heart muscle (B2) and myotubes differentiated from myoblasts *in vitro* (B3). Note that only muscle nuclei (solid arrowheads) exhibit TLs. In muscle fibroblasts (arrows) or undifferentiated cultured myoblasts (empty arrowheads), *Ttn* is condensed at the nuclear periphery.

C, The *Neb* gene is expressed in skeletal muscles and cultured myotubes, although to a lesser degree than *Ttn*. Accordingly, it forms smaller TLs. Arrowheads indicate muscle nuclei; arrows indicate fibroblast nuclei with silent *Neb*.

D, E, The *Myh11* (D) and *Cald1* (E) genes are expressed in smooth muscles of colon and bladder and form TLs in the nuclear interior of these cells. Note that after RNA-FISH, only smooth muscles (*arrowheads*) but not the neighboring epithelial cells (*arrows*) exhibit TLs. In addition, *Cald1* is expressed in cultured myoblasts and forms small TLs in these cells.

Most of the panels display FISH with simultaneous detection of DNA and RNA signals; panels with only RNA-FISH are marked. All images are *projections of 1-3 µm* confocal stacks. Scale bars for overviews of skeletal muscle, colon and bladder, *50 µm*; for rest of the panels, *5 µm*.

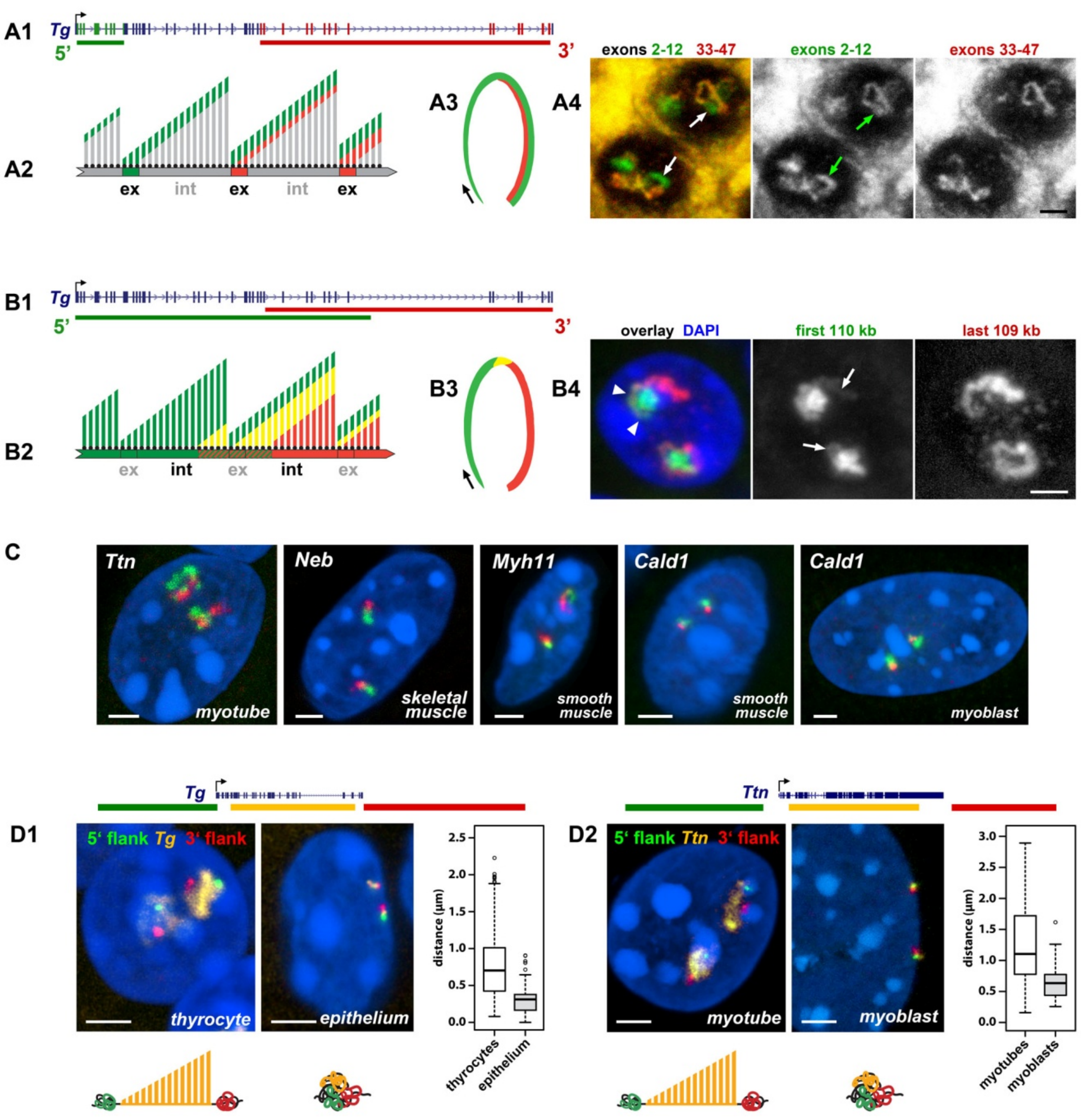

A1
Tg
5'
3'
A2
ex
int
ex
int
ex
A3
A4
exons 2-12 33-47
exons 2-12
exons 33-47
B1
Tg
5'
3'
B2
ex
int
ex
int
ex
B3
B4
overlay DAPI
first 110 kb
last 109 kb
C
Ttn
myotube
Neb
skeletal muscle
Myh11
smooth muscle
Cald1
smooth muscle
Cald1
myoblast
D1
Tg
5' flank Tg 3' flank
thyrocyte
epithelium
distance (µm)
2.5
2.0
1.5
1.0
0.5
0.0
thyrocytes
epithelium
D2
Ttn
5' flank Ttn 3' flank
myotube
myoblast
distance (µm)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
myotubes
myoblasts

Figure 2. TLs manifest transcription progression, including co-transcriptional splicing, and dynamically modify harboring chromosomal loci

A, Successive labeling of *Tg* TLs with probes for 5' and 3' exons. A1, Distribution of 11 *5'* exons (*green*) and 15 *3'* exons (*red*) of the *Tg* gene chosen for RNA-FISH. A2,3, Schematics showing changes in the composition of nRNAs during transcription progression: nRNAs of the first half of a TL contain only "*green*" 5' exons; in the second half they also include "*red*" 3' exons (A2). Correspondingly, "*green*" exons label transcripts along the whole TL and "*red*" exons label only transcripts of the TL second half (A3). A4, Examples of *Tg* TLs labeling after RNA-FISH with probes described in A1. Probe for the 5' exons labels the entire TL, whereas the probe for 3' exons labels second half of the TL. *Arrows* point at the TL regions labeled with only 5' probe.

B, Sequential labeling of *Tg* TLs with genomic probes highlighting introns. B1, coverage of the *Tg* gene with two overlapping BACs, for 5' (*green*) and 3' (*red*) gene halves. B2,3, Schematics explaining changes in nRNA labeling with probes highlighting introns during transcription progression: "*green*" 5' introns are gradually spliced out and replaced by "*red*" 3' introns (B2). Correspondingly, "*green*" introns label transcripts of the first and "*red*" exons of the second TL halves with some overlap (*yellow*, B3). B4, Example of *Tg* TLs after RNA-FISH with two BAC probes described in B1. BAC probes highlighting 5' (*green*) and 3' (*red*) introns sequentially label the TLs as a result of co-transcriptional splicing. In addition, since the 5' probe includes 5' exons, it also faintly labeled the second half of the loop (*arrows*). The region hybridized by both overlapping BACs is marked with *arrowheads*.

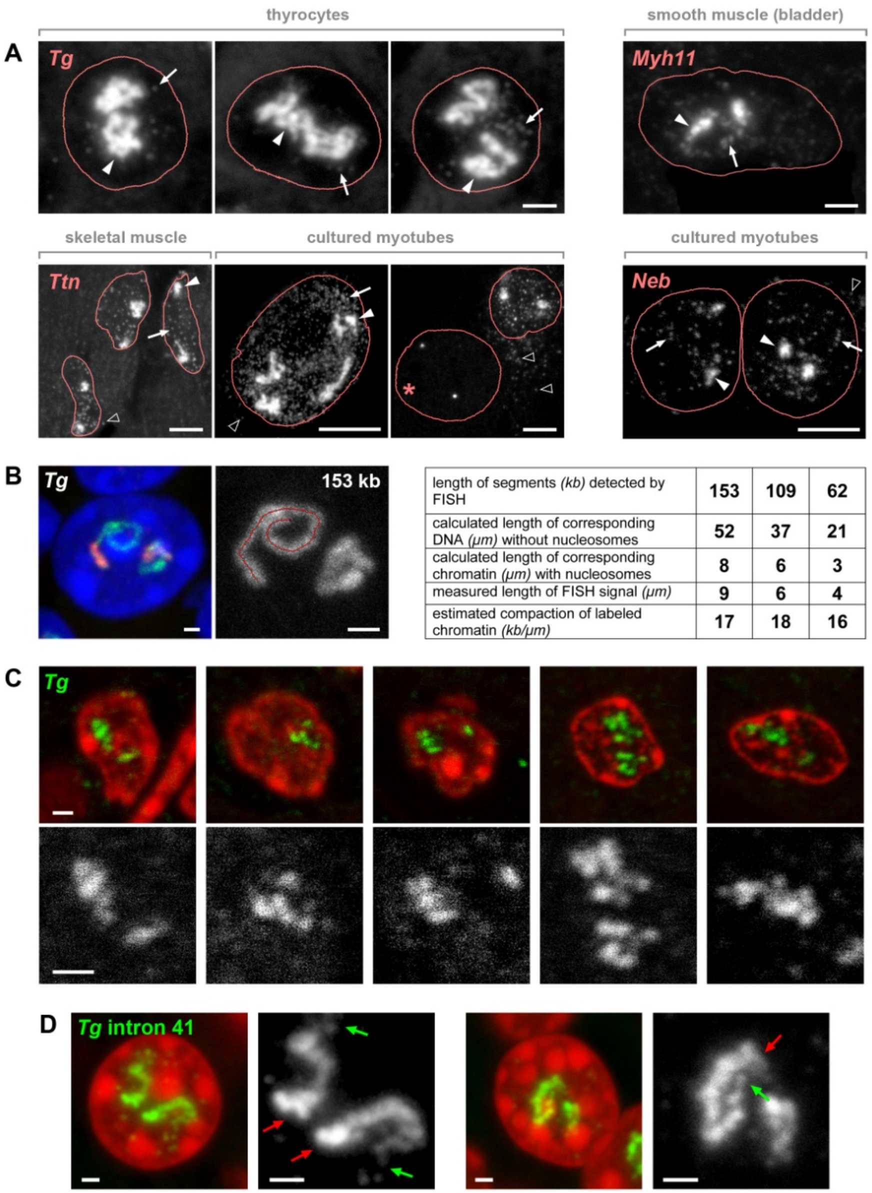
ex, exons; *int*, introns. *Black arrows* indicate direction of transcription. Scale bars, *2 µm*

C, TLs formed by other long highly expressed genes exhibit co-transcriptional splicing. RNA-FISH with BAC probes encompassing 5' (*green*) and 3' (*red*) regions of the genes (see SI Table S3 for the probes).

Projections of confocal sections through 1 - 2 *µm*. Scale bars, *2 µm*

D, Expressed genes expand from the harboring loci and separate their flanks. Distances between 5' (*green*) and 3' (*red*) flanking regions of the *Tg* (D1) and *Ttn* (D2) genes are larger in cells expressing (*left panels*) compared to control cells not expressing the genes (*right panels*), as shown by schematics at the bottom. Boxplots depicting the 3D distance between the flanking regions in expressing and not-expressing cells are shown on the right. The median inter-flank distance for *Tg* in thyrocytes is 2.3-fold larger than in control epithelial cells (703 nm *versus* 311 nm). The median inter-flank distance for *Ttn* in myotubes is 1.7-fold larger than in control myoblasts (1,104 nm *versus* 634 nm).

FISH with simultaneous detection of DNA and RNA signals. Projections of confocal sections through 1 - 3 *µm*. Scale bars, *2 µm*



length of segments *(kb)* detected by FISH	153	109	62
calculated length of corresponding DNA *(µm)* without nucleosomes	52	37	21
calculated length of corresponding chromatin *(µm)* with nucleosomes	8	6	3
measured length of FISH signal *(µm)*	9	6	4
estimated compaction of labeled chromatin *(kb/µm)*	17	18	16

Figure S3 (related to Figure 2). Nucleoplasmic granules in cells with highly expressed long genes and TL compaction level

A, Genomic probes used for visualization of TLs (*arrowheads*) hybridize to numerous nucleoplasmic granules (*arrows*) in thyrocytes and muscle cells. In case of *Tg* and *Myh11*, the granules most likely represent accumulations of excised but not yet utilized introns. In case of *Ttn* and *Neb* with exceptionally long mRNAs (ca.101 kb and 22 kb, respectively), at least, a proportion of granules might represent mature mRNAs. In agreement with this, RNA-FISH detects similar granules in the myotube cytoplasm (*empty arrowheads*).

For clarity, DNA counterstain is omitted and only *contours* of the nuclear projections are outlined. *Asterisk* marks the nucleus of a myoblast without *Ttn* TLs and granules. Projections of confocal sections through 1.5 – 2.5 *µm*. Scale bars: thyrocytes and smooth muscle, 2 *µm*; skeletal muscle and cultured myotubes, 5 *µm*.

B, To assess the compaction level of TLs, the contour length of the *Tg* TL regions was evaluated on projections of thyrocyte nuclei after RNA-FISH. The track of the *Segmented Line tool* in ImageJ, used for measurements, is shown on the right panel. Scale bars: 1 *µm*.

Tg regions of 153 kb, 109 kb and 62 kb had a similar compaction level and measured 9 µm, 6 µm and 4 µm, respectively. These values rather correspond to a nucleosomal structure of TL chromatin (*table*). However, since the measurements were performed on maximum intensity projections, which mask curving and folding of TLs, the length of measured segments is rather underestimated.

C, TLs exhibit fine coiling and folding not fully resolvable by confocal microscopy but detected on thin resin (*Lowicryl*) sections with thickness of 50-70 nm. The upper panels include representative images of thyrocyte nuclei after RNA-FISH. The lower panels show 2-fold close ups of the corresponding *Tg* TLs images, exhibiting curling and twisting.

Single optical sections. Scale bars: 1 *µm*.

D, Coiling of *Tg* intron 41 (54 kb): the left panels show two representative images of thyrocyte nuclei after RNA-FISH detecting only the intron; the right panels show 2-fold close ups of the introns, which exhibit curling and spiraling. Note a thickness gradient of intron signals reflecting growth of nRNA transcripts from 5’ intron beginning (*green arrows*) towards 3’ intron end (*red arrows*).

Projections of confocal sections through 2 *µm*. Scale bars: 1 *µm*.

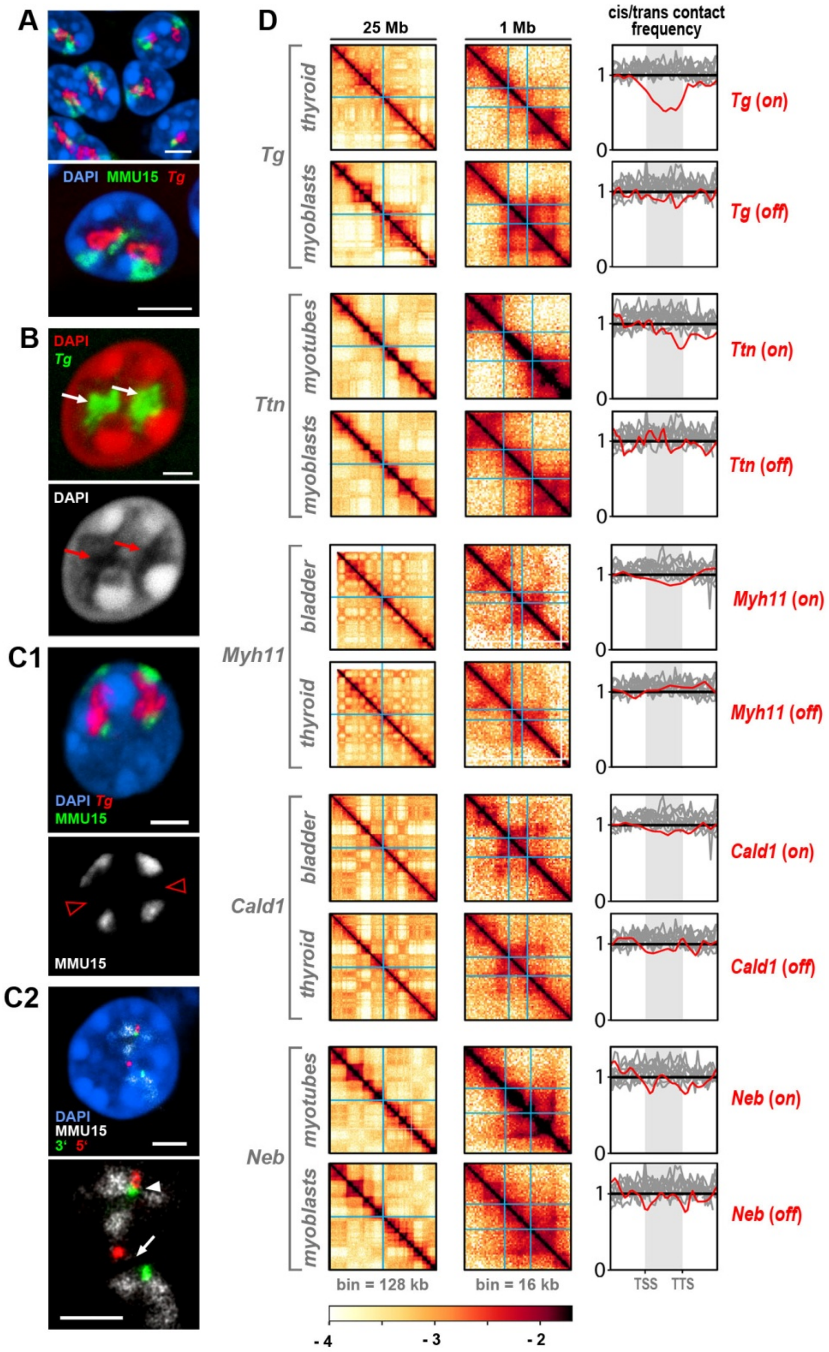

A
DAPI MMU15 Tg
B
DAPI
Tg
DAPI
C1
DAPI Tg
MMU15
MMU15
C2
DAPI
MMU15
3' 5'
D
25 Mb
1 Mb
cis/trans contact frequency
Tg
thyroid
myoblasts
Tg (on)
Tg (off)
Ttn
myotubes
myoblasts
Ttn (on)
Ttn (off)
Myh11
bladder
thyroid
Myh11 (on)
Myh11 (off)
Cald1
bladder
thyroid
Cald1 (on)
Cald1 (off)
Neb
myotubes
myoblasts
Neb (on)
Neb (off)
1
0
bin = 128 kb
bin = 16 kb
TSS
TTS
- 4
- 3
- 2

Figure 3. Transcription Loops are excluded from harboring chromosomes.

A, *Tg* TLs (*red*) emanate from harboring chromosome 15 (*green*) and protrude into the nuclear interior.

B, Nucleoplasmic regions corresponding to *Tg* TLs (*white arrows*) are depleted of chromatin (*red arrows*).

C1, Ca. 2% of *Tg* alleles (*red*) split chromosome territories (*green*) into two halves with an unpainted gap between them (*red empty arrowheads in the lower panel*).

C2, The gap in the chromosome territory (*arrow*) is marked with the 5' (*red*) and 3' (*green*) *Tg* flanking sequences. The second homologue is not split; the 5' and 3' *Tg* flanks are in close proximity (*white arrowhead*).
Projections of confocal stacks through 0.5 – 1 *µm*; scale bars: A, 5 *µm*; B-C, 2 *µm*.

D, For selected genes, a 25 Mb Hi-C contact map and a 1 Mb zoom view are shown both for a cell type where the gene is expressed (*on*) and for one where it is silent (*off*). TSS and TTS of the genes are marked with light blue lines. The rightmost column shows *cis/trans* ratio profiles, i.e. the total number of Hi-C contacts of a locus with loci on the same chromosome divided by the total number of contacts with other chromosomes, calculated near gene of interest (*red*) and compared to 9-10 other long genes with low expression (*gray*, see SI TableS2 for gene lists). For comparability of *cis/trans* ratio profiles, the x-coordinates are rescaled such that TSS and TTS of all genes in a panel align (the *gene bodies are marked by shaded areas*). To highlight potential dips localized in the gene body against longer range variations, *cis/trans* ratio profiles are normalized to unity in the region outside the gene body. The *cis/trans* ratio for the *Tg* gene shows a pronounced dip in thyroid; *Ttn* in myotubes and *Myh11* in bladder show a moderate dip; profiles for *Cald1* and *Neb* are not significantly different from those in not expressing cells or other long lowly expressed genes.

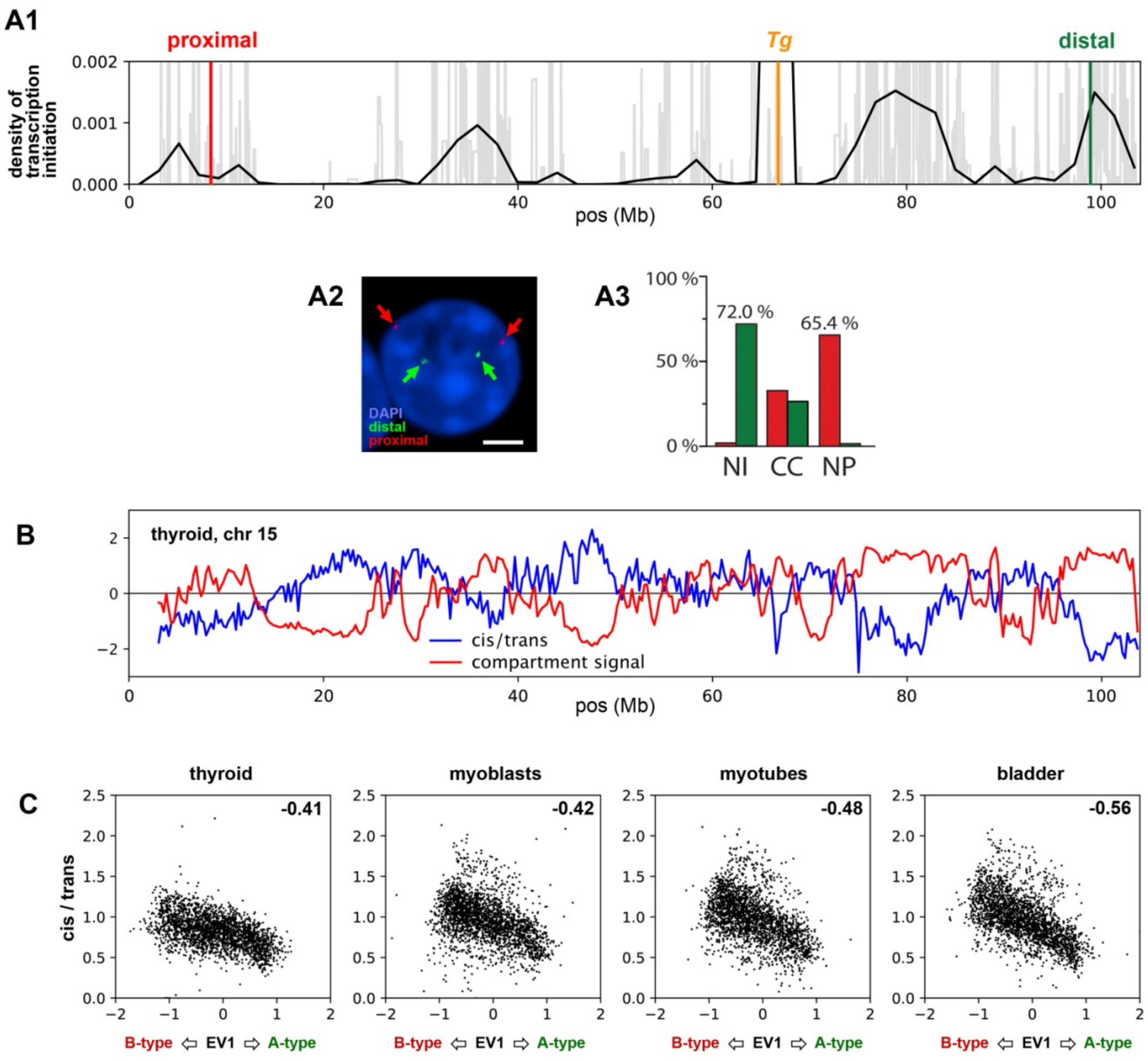


Figure S4 (related to Figure 3). Asymmetry of chromosome 15 and enrichment of loci with low *cis/trans* ratios within A compartments

A1, RNA-seq analysis of thyrocytes reveals an asymmetrical distribution of expressed loci (*light-gray*) along chromosome 15: the majority of expressed loci are concentrated towards the *distal end* from the centromere region. The extremely highly upregulated *Tg* (*orange line*) is positioned between regions with high expression (*to the right*) and lower expression (*to the left*). Positions of the BAC probes for proximal and distal chromosome ends for the FISH experiment (A2) are shown as *red* and *green lines*, respectively.

A2, Polarized positioning of chromosome 15, detected using BAC probes for proximal and distal chromosome ends, with the proximal end (*red*) positioned mostly at the nuclear periphery and the distal end (*green*) protruding into the nuclear interior. Projection of a confocal stack through 1.5 *µm*; scale bar, *2 µm*.

A3, Quantification of FISH signals confirmed preferential positioning of chromosome 15 ends: distal signals (*green*) are found in the permissive compartment of the nuclear interior (*NI*), whereas proximal signals (*red*) reside in the silencing compartments, such as the nuclear periphery (*NP*). Both proximal and distal signals were adjacent to chromocenters (*CC*) in equal proportions.

B, The compartment profile, shown by a *red line* and corresponding to A (high EV values) and B (low EV values), and profile of *cis/trans* ratio, shown by a *blue line*, negatively correlate along chromosome 15. For clarity, both tracks were transformed by subtracting the chromosomal mean and setting the standard deviation to unity.

C, *Cis/trans* ratios are lower in A than in B compartments for all four studied cell types. The scatter plots are computed from the genome wide compartment profile and the *cis/trans* ratio profile, both at a bin size of 1,024 kb. The Pearson correlation coefficients are indicated in the upper right corners of the scatter plots.

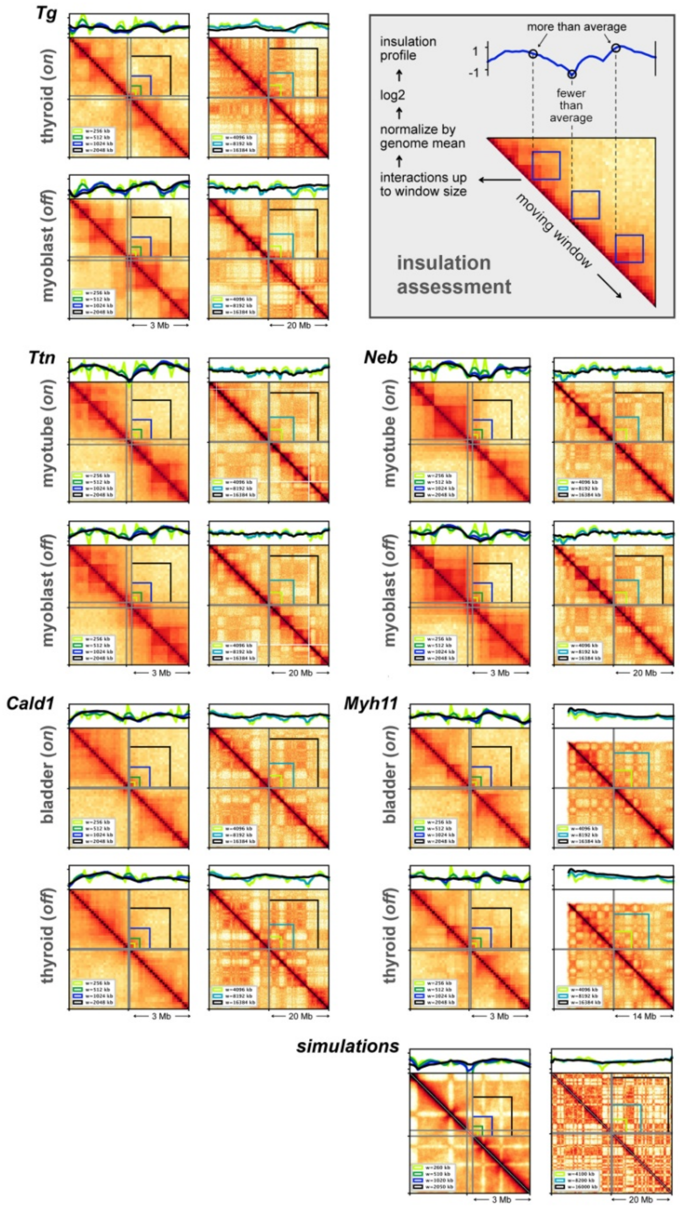


Figure S5 (related to Figure 3). TLs do not cause insulation at different length scales.

Insulation (Crane et al. 2015) assesses Hi-C contacts spanning across a given locus up to a maximal distance *w* (see top right insert). Contacts in a square window of size *w* were aggregated and the square was slid along the Hi-C diagonal. The score was normalized by its genome wide mean. Profiles show log2 of the score, such that a locus with profile value -1 has a two-fold reduced number of contacts spanning the locus up to distance *w* compared to the genome wide mean. Insulation scores are computed with the *cooltools* package (https://github.com/mirnylab/cooltools). We computed insulation profiles for Hi-C maps with bin size of 128 kb for various window sizes from 256 kb up to ≈16 Mb.

For every analyzed gene, the *left* and *right columns* show a 3 and 20 Mb Hi-C map with insulation profiles for different window sizes; the *top* and *bottom panels* show insulation profiles in expressing (*on*) and not-expressing (*off*) cells, respectively.

The analysis shows little correlation between insulation and the formation of TLs: insulation profiles at the gene loci do not differ much between cell types with the gene *on* or *off*. For example, the *Tg* gene shows a moderate dip at scales up to ~≈ 1Mb in both thyroids (*on*) and myoblasts (*off*), and no dip in either cell type on the larger scale.

Analysis of simulated TLs (see Fig.7) confirmed that TL formation does not cause large scale insulation (*bottom*).

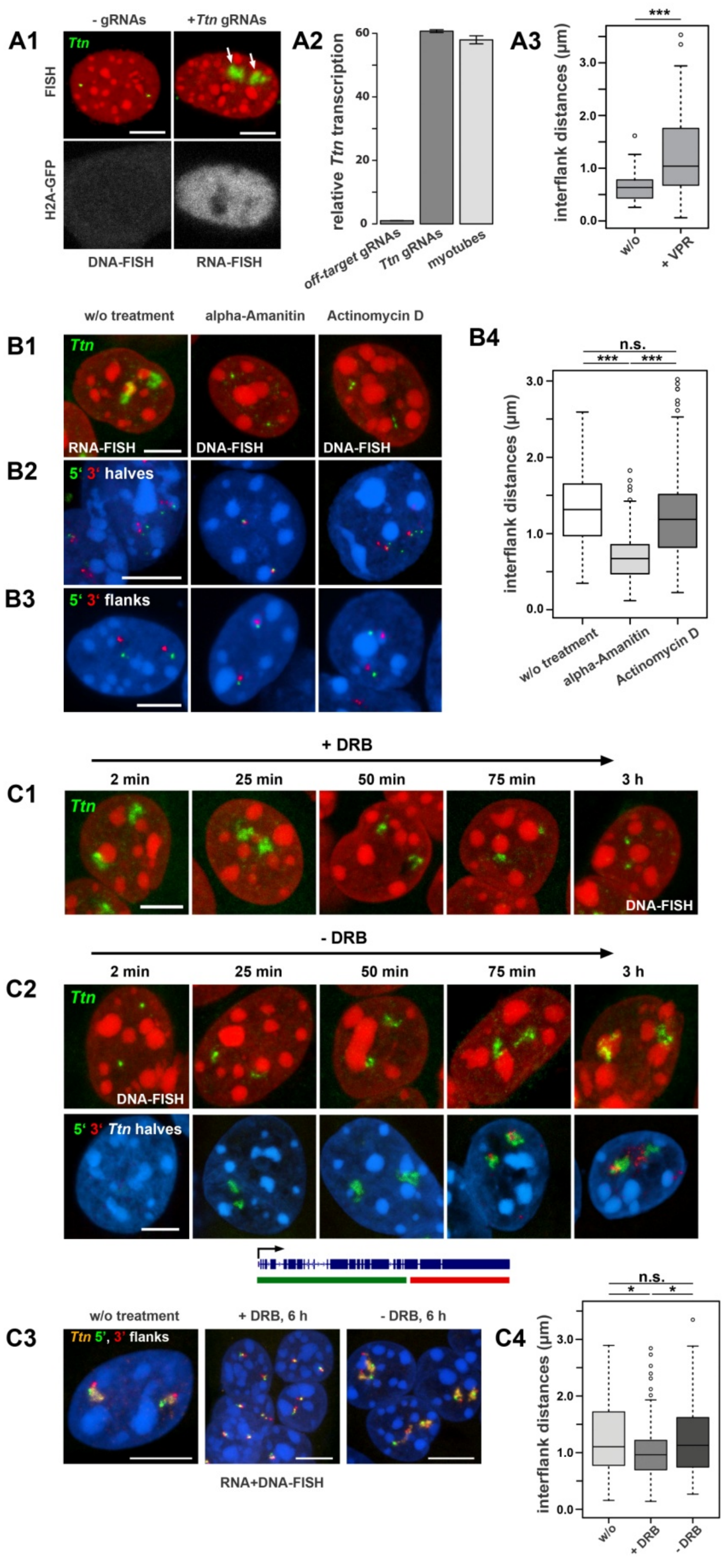

A1
- gRNAs
+Ttn gRNAs
Ttn
FISH
H2A-GFP
DNA-FISH
RNA-FISH
A2
relative Ttn transcription
60
40
20
0
off-target gRNAs
Ttn gRNAs
myotubes
A3
***
interflank distances (µm)
3.0
2.0
1.0
0.0
w/o
+ VPR
w/o treatment
alpha-Amanitin
Actinomycin D
B1
Ttn
RNA-FISH
DNA-FISH
DNA-FISH
B2
5' 3' halves
B3
5' 3' flanks
B4
n.s.
***
***
interflank distances (µm)
3.0
2.0
1.0
0.0
w/o treatment
alpha-Amanitin
Actinomycin D
+ DRB
2 min
25 min
50 min
75 min
3 h
C1
Ttn
DNA-FISH
- DRB
2 min
25 min
50 min
75 min
3 h
C2
Ttn
DNA-FISH
5' 3' Ttn halves
w/o treatment
+ DRB, 6 h
- DRB, 6 h
C3
Ttn 5', 3' flanks
RNA+DNA-FISH
C4
n.s.
*
*
interflank distances (µm)
3.0
2.0
1.0
0.0
w/o
+ DRB
- DRB

Figure 4. TLs are formed or obliterated in response to transcription induction or inhibition

A, Ectopic induction of *Ttn* TLs in myoblasts not expressing *Ttn.* Myoblasts, stably expressing dCas9 conjugated with VPR, were co-transfected with plasmids for H2A-GFP and with either gRNAs targeting the *Ttn* promoter or with off-target gRNAs. As assessed by qPCR, *Ttn* expression was induced in cells transfected with *Ttn* gRNAs to a level comparable to untreated myotubes and exceeding *Ttn* expression in control cells transfected with off-target gRNAs by ca. 60-fold (A2). *Error bars are standard deviations.* Accordingly, 90% of transfected myoblasts exhibited *Ttn* TLs (A1, *right*) and increased 2-fold median inter-flank *Ttn* distances (from 634 nm to 1,039 nm) (A3). In untransfected cells, the *Ttn* gene remained condensed (A1, *left*) with converged flanks (A3). ****p-value < 0.001, Wilcoxon rank sum test.*

B, Treatment with transcription inhibitors eliminates *Ttn* TLs in differentiated myotubes. B1, *Ttn* TLs are detected in untreated cells (*left*) but not in cells treated with inhibitory drugs, α-amanitin (*mid*) and actinomycin D (*right*). B2,3, According to different mechanisms of transcription inhibition, the two drugs affect gene condensation in different ways. As DNA-FISH with probes for 5' and 3' halves (B2) and 5' and 3' flanks (B3) shows, α-amanitin treatment caused strong condensation of the gene body and convergence of the flanks, whereas after actinomycin D treatment, the gene body remains decondensed and flanks remain diverged, similar to *Ttn* in untreated cells (B4). ****p-value < 0.001, n.s. = not significant, Wilcoxon rank sum test.*

C, Application of DRB, a reversible transcription inhibitor, causes inhibition of transcription elongation and thus leads to gradual shrinkage of *Ttn* TLs (C1); removal of DRB revives elongation and leads to a gradual restoration of *Ttn* TLs (C2, *top*). Differential labeling of 5' (*green*) and 3' (*red*) halves of Ttn (*see gene labeling scheme*) allows better monitoring the TLs dynamics (C2, *bottom*): the signal for the 5' end of TL appears first, the signal for 3' end appears only after ca. 1 h. In correspondence to TL dynamics, median distances between *Ttn* flanks are decreased after complete transcription inhibition from 1,104 nm to 963 nm, but remain larger than inter-flank distances in myoblasts (634 nm), and then restored up to 1130 nm upon transcription restoration (C3,4). **p-value < 0.05, n.s. = not significant, Wilcoxon rank sum test.* For used *pseudo-colors* and more inter-flank distances see Figure 2D2. Panels in C1 and C2, if not indicated otherwise, show RNA-FISH. Panels in C3 show FISH detecting both RNA and DNA.

For all panels: Projections of confocal sections through 2-5 *µm*. Scale bars, 5 *µm*.

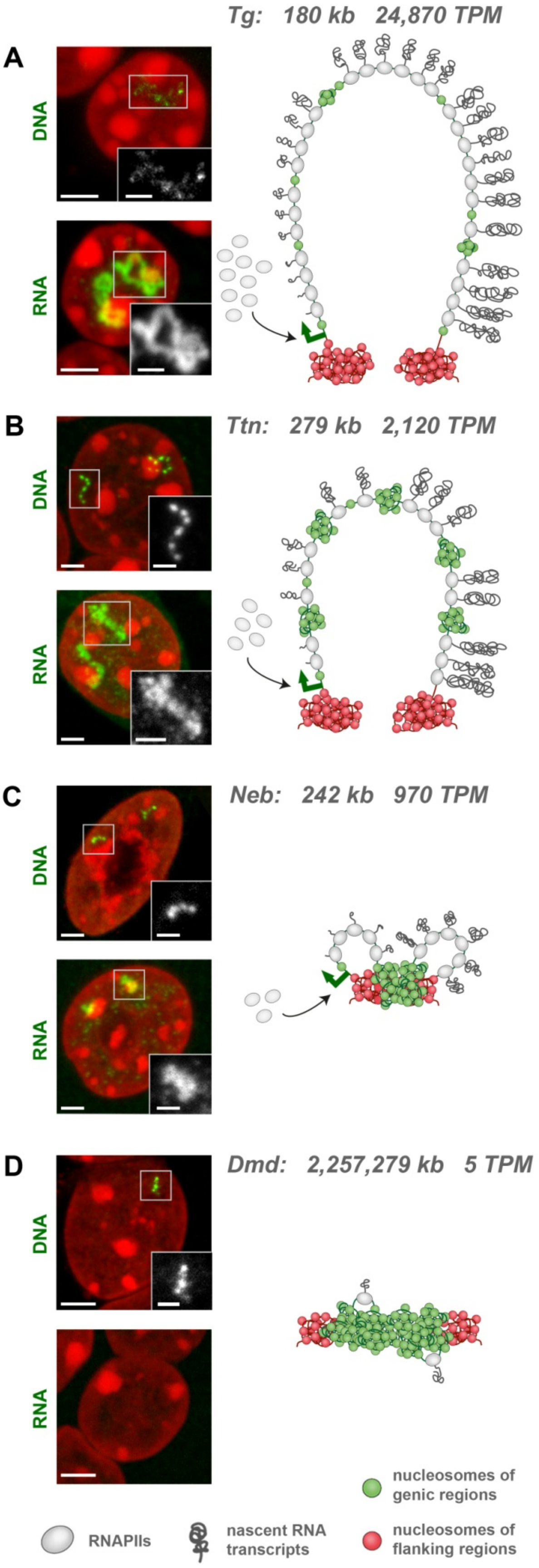

Tg: 180 kb 24,870 TPM
A
DNA
RNA
B
Ttn: 279 kb 2,120 TPM
DNA
RNA
C
Neb: 242 kb 970 TPM
DNA
RNA
D
Dmd: 2,257,279 kb 5 TPM
DNA
RNA
nucleosomes of genic regions
RNAPIIs
nascent RNA transcripts
nucleosomes of flanking regions

Figure 5. Level of gene expression is manifested by a pattern of microscopically detectable transcription bursts and pauses

Examples of four long genes - *Tg* (**A**), *Ttn* (**B**), *Neb* (**C**) and *Dmd* (**D**) - are arranged from top to bottom according to their expression level in the respective cell type. Length (kb) and expression level (TPM) are indicated next to the gene names. For every gene two images are displayed *on the left*, after DNA-FISH detecting the gene body and after RNA-FISH detecting nRNA transcripts. The more a gene is expressed, the less solid the DNA-signal is and the more expanded the RNA-signal is. *Vice versa*, the less a gene is expressed, the more condensed the gene body is and the less extended the RNA-signal is. The schematics *on the right* are interpretations of the observed FISH signal patterns in terms of transcriptional bursts, depicted as RNAPII convoys with attached nRNA transcripts, and transcriptional pauses, depicted as condensed chromatin (*green nucleosomes*). For simplicity, splicing events are not depicted on the schemes. For more examples of DNA- and RNA-FISH for *Tg, Ttn* and *Neb*, see Fig.1C,D.

Projections of 1-1.5 *µm* confocal stacks. Scale bars, *2 µm*; in insertions, *1 µm*.

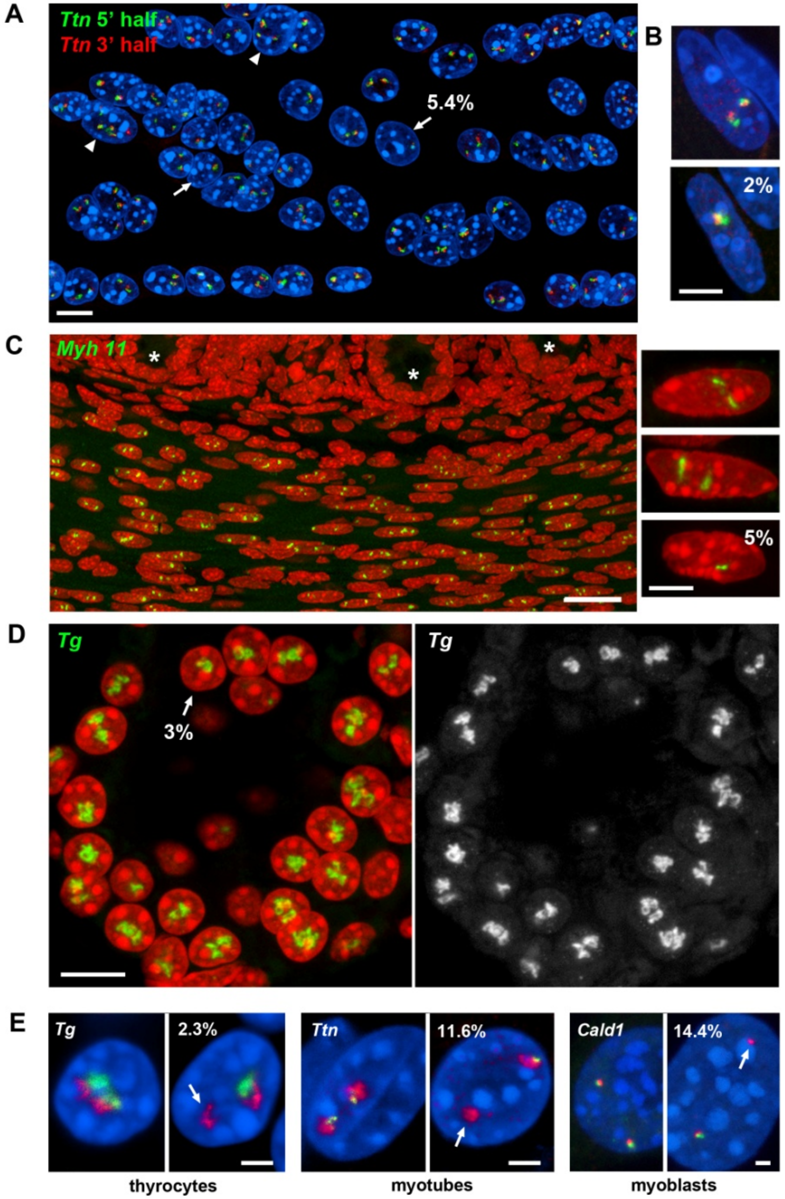


Figure S6 (related to Figure 5). Highly expressed genes are expressed biallelically and exhibit transcriptional bursting.

A-D, Examples of biallelic and monoallelic expression of the *Ttn* gene in cultured myotubes (A) and skeletal muscle (B), of the *Myh11* gene in smooth muscles of intestine (C) and of the *Tg* gene in thyroid gland (D) as detected by RNA-FISH. Proportion of nuclei with monoallelic expression (%) is indicated. *Arrows* point at the nuclei with one transcribed allele; *arrowheads* on (A) point at tetraploid nuclei (4 signals); *asterisks* in (C) mark intestinal crypts; note that epithelial cells and fibroblasts do not exhibit *Myh11* RNA-FISH signal.

E, RNA-FISH with oligo probes (*green*) for 30 kb regions of the *Tg*, *Ttn* and *Cald1* reveals that a proportion of TLs does not exhibit a FISH signal (*arrows*), indicating that the respective 30 kb region is not expressed at this particular time and in these particular alleles. To ensure that both alleles were expressed in the studied cells, the rest of the TLs were co-labeled with genomic probes (*red*). The proportion of alleles without the 30 kb probe signal is lowest for *Tg* (2.3%) in comparison to *Ttn* (11.6%) and *Cald1* (14.4%), indicating that in lower expressed genes the 30 kb region more often falls into a transcription pause than in highly expressed genes, which in turn means that transcription pauses are longer in these genes.
Projections of confocal stacks through 2-4 *µm*. Scale bars: A,D, 10 *µm*; B, C (*insertion*), 5 *µm*; C (*overview*), 50 *µm*; E, 2 *µm*

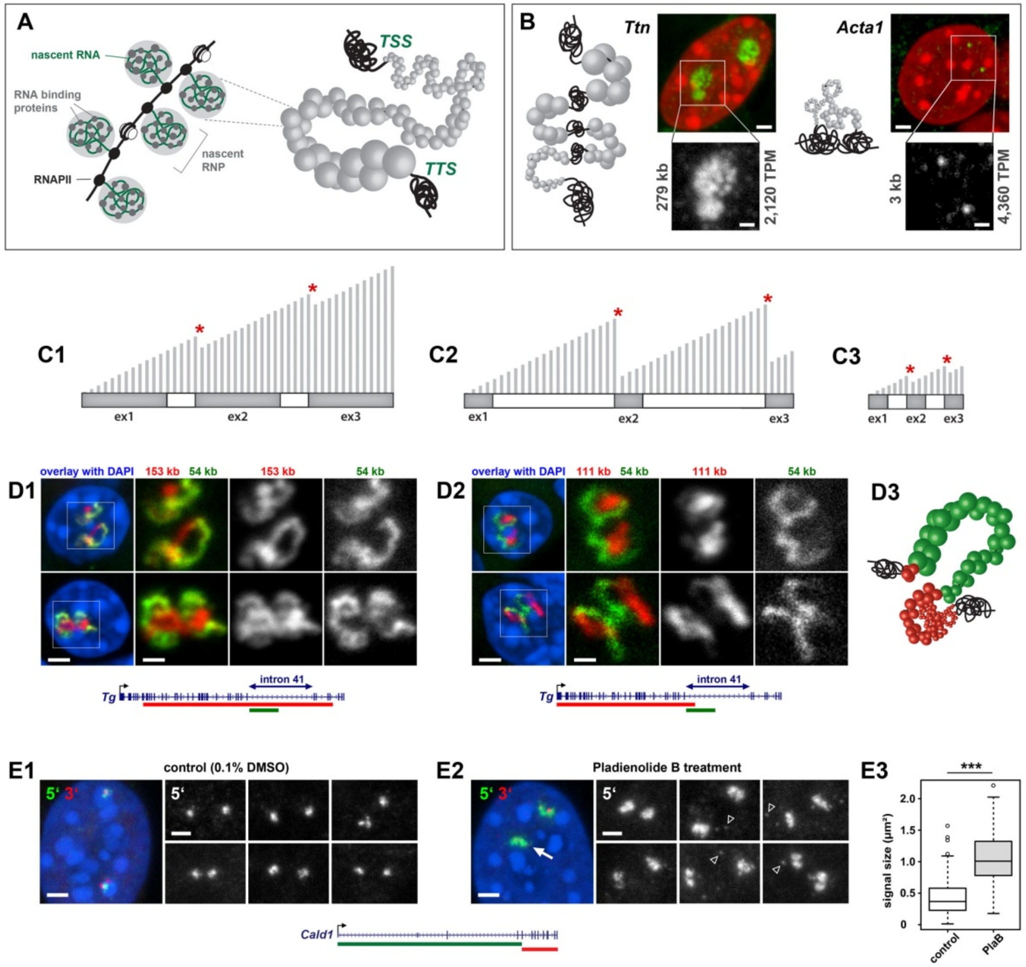


Figure 6. Transcription loop expansion from the harboring locus can be explained by increased stiffness of the gene axis

A, Schematics depicting the formation of voluminous nRNPs consisting of nRNAs bound by multiple RNA binding proteins. In case of long genes, especially genes with long introns, nRNPs decorating gene body form voluminous structures. Due to high expression, a gene axis becomes densely decorated by multiple voluminous nRNPs, loses its flexibility and thus acquires stiffness forcing the gene to protrude into the nuclear space.
B, Comparison of TLs formed by *Ttn* and *Acta1* in myotube nuclei. The *Ttn* mRNA is ca. 101 kb and therefore the gene axis is decorated by nRNPs progressively growing in size, which results in extension of the *Ttn* TLs. In striking contrast, the *Acta1* gene is only 3 kb and mRNA transcribed from *Acta1* is only 1.7 kb. Therefore, the gene axis is decorated by small nRNPs, remains flexible and coils, even though *Acta1* is highly expressed. Schematics explaining the difference in TL morphologies are shown on the *left*; RNA-FISH images are shown on the *right*. Projections of confocal stacks through 0.9 – 1.2 *µm*. Scale bars: RGB panels, 2 *µm*, close ups, 1 *µm*.
C, Schematics showing the length of nRNAs build on long gene with short introns (C1), long gene with long introns (C2) and short gene consisting of short exons and introns (C3). Exons and introns are shown as *dark-grey* and *white rectangles*, respectively; transcripts are depicted as perpendicular *light-grey lines*; due to space limitation, the transcripts depicted with length of only half of the template length; splicing events are marked with *red asterisks*. The nRNA lengths on C1 and C2 are similar, but nRNAs length on C3 is incomparably smaller.
D, Disproportional extension of *Tg* intron 41 (54 kb) in comparison to the mid part of the gene (153 kb, D1) or to 5' half of the gene (111 kb, D2). In both cases, the *Tg* TL region encompassing the intron (*green*) is extended stronger than other TL regions (*red*), presumably as a consequence of increased stiffness of the large intronic part decorated by nRNAs growing from ca. 6 to ca. 60 kb. Projections of 1.2 *µm* confocal stacks. Scale bars: panels with entire nuclei, 2 *µm*, close ups, 1 *µm*. The *Tg* gene structure and probe coverage is shown at the *bottom*. D3, Schematics showing protrusion of *Tg* TL region carrying voluminous RNPs formed over intron 41 (*green*) in comparison to coiled 5' and 3' TL regions carrying smaller nRNPs (*red*).
E, Comparison of *Cald1* TL size in control myoblasts (0.1% DMSO, 4 h; E1) and myoblasts treated with splicing inhibitor Pladienolide B (10nM, 4 h; E2). RNA-FISH signals of 5' and 3' ends are shown in *green* and *red*, respectively. RGB images are supplemented with representative gray-scale images of the 5' gene end, which includes long introns. Projections of 3-5 *µm* confocal stacks. Scale bars: 2 *µm*. Scheme at the *bottom* shows the gene structure and probe coverage. Both phenomena, absence of 3' signal in a large proportion of *Cald1* alleles (*arrow*) and accumulation of nucleoplasmic granules detected by 5' and 3' probes (*arrowheads*), indicate that Pladienolide B treatment causes abortion of transcription. Nevertheless, the RNA-FISH signals of the 5' gene end are increased 2.5 - fold after splicing inhibition (E3). Measurements were performed on projections of confocal stacks. ****p-value < 0.001, Wilcoxon rank sum test.*

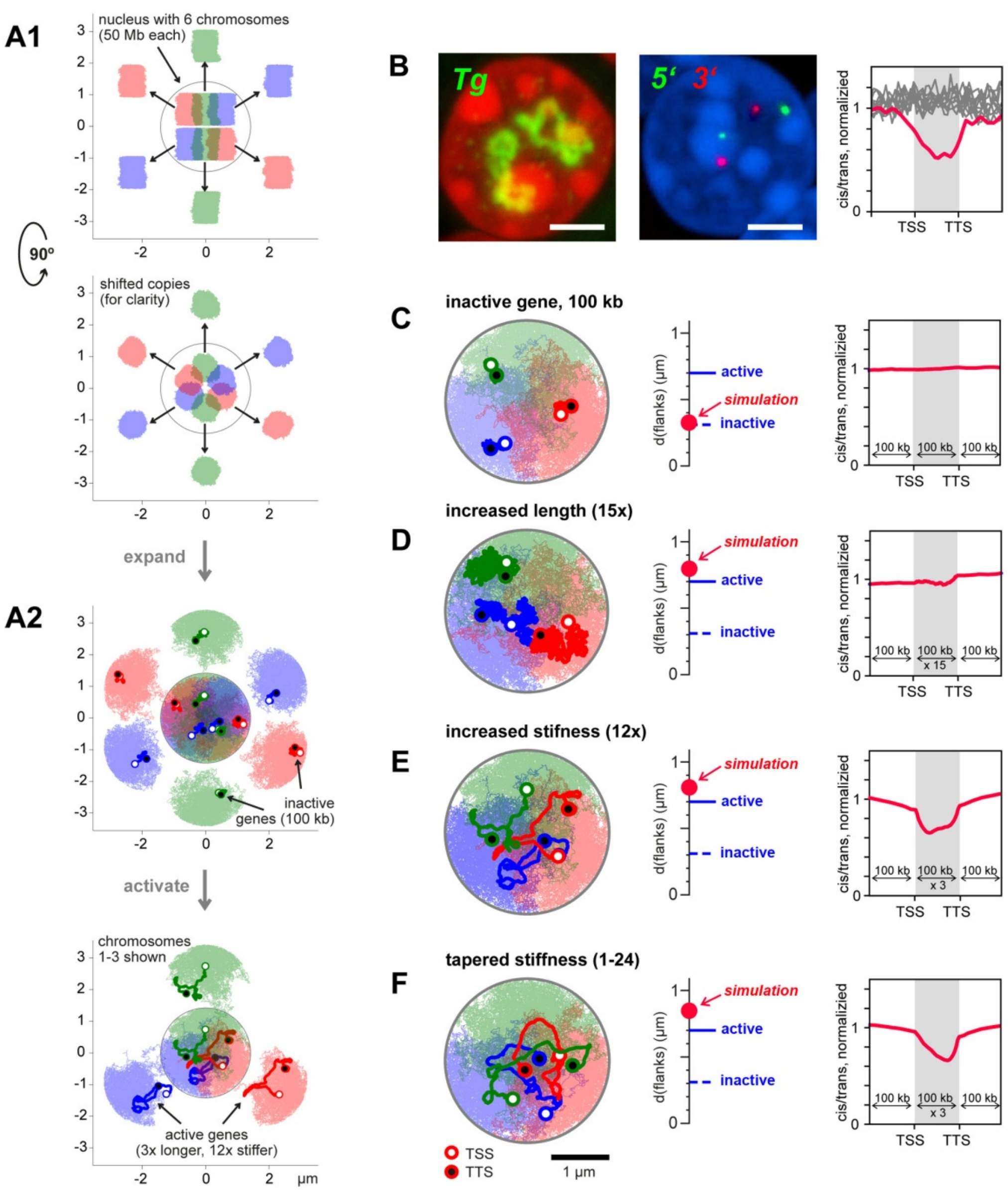


Figure 7. Polymer modeling shows that increased stiffness of transcribed genes can explain gene expansion, increased inter-flank distance and reduced *cis/trans* contact ratio.

A, Polymer simulation setup. **A1**, Six territorial chromosomes (50 Mb each) in a spherical nucleus were obtained by initiating them in a mitotic-like state (top and side view). For clarity, shifted copies are shown outside the nucleus. **A2**, polymers were expanded and for each polymer, a gene of interest of 100 kb was assigned (depicted with darker color). For clarity, shifted copies of six chromosomes with simulated genes are shown separately (top). In the rest of simulations only 3 chromosomes are shown (bottom).

B, Biological variables used to verify TL models: appearance of *Tg* TLs, inter-flank distances and *cis/trans* contact ratio calculated from Hi-C maps.

C, Genes of 100 kb length in the inactive state (no increased stiffness) are compact structures and don't exhibit a dip in *cis/trans* contact ratio compared to the surrounding fiber.

D, A mere increase in gene contour length (15-fold) leads to a bigger TL volume but doesn't exhibit a clearly discernible contour or a dip in *cis/trans* contact ratio.

E, Increased stiffness (12-fold) together with a moderate increase in contour length (3-fold) shows that TLs expand substantially, exhibit a clear contour, and a dip in *cis/trans* contact ratio.

F, A gradual increase in stiffness along the gene (in average 12-fold) leads to larger flank separation, more coiling of the 5' compared to 3' gene ends and an asymmetric dip in *cis/trans* contact ratio with a steeper slope at the 3' end.

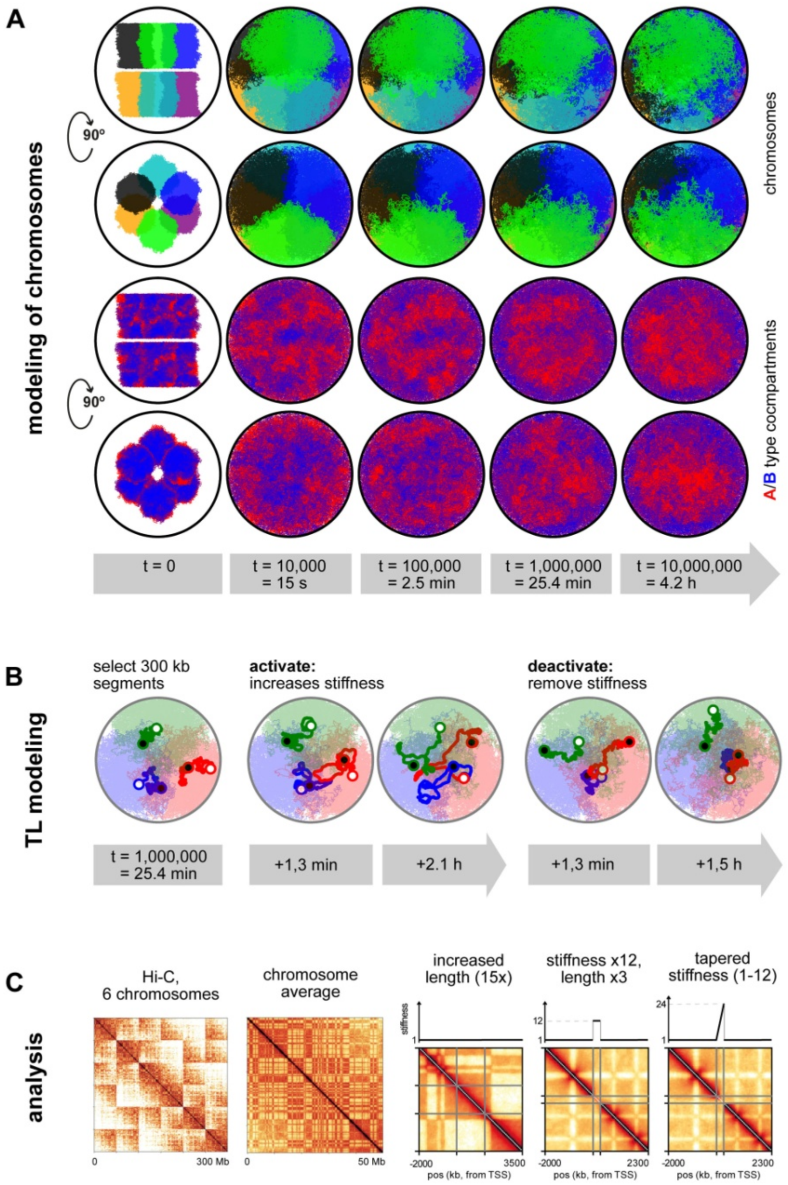


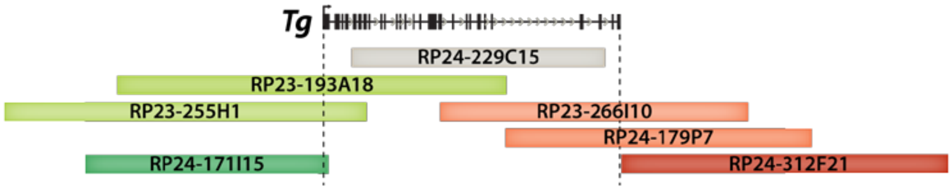
Figure S7 (related to Figure 7). Polymer simulation of chromosomes.

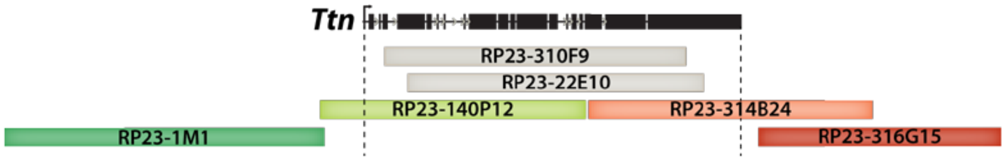
A, Six chromosomes (50 Mb each) were initiated in a mitotic-like state with unit volume density. Row 1 and 3 show top views, row 2 and 4 show side views. In rows 1 and 2 six chromosomes are differentially colored; in rows 3 and 4 compartmental segments of A and B type chromatin are differentially colored with red for A and blue for B compartments. The initial expansion is very fast (column 2). However, once the chromosomes fill the nucleus uniformly, subsequent dynamics is very slow and chromosomes retain their territoriality (note that times increase logarithmically). Nevertheless, due to attraction of B-type chromatin to the lamina, a radial structure emerges (rows 3 and 4).

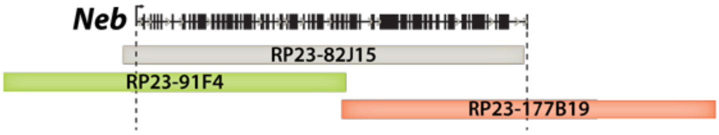
B, Transcription loops are modeled by choosing a 300 kb segment on each chromosome 25.4 minutes after expansion and increasing the stiffness of of the polymer fiber. The genes quickly expand on the order of minutes and were simulated for approximately 1.5 h. The measurements of inter-flank distances and Hi-C maps were performed using configurations sampled from the second half of this time interval. When genes are deactivated by removing the excess stiffness, they collapse back to the inactive state.

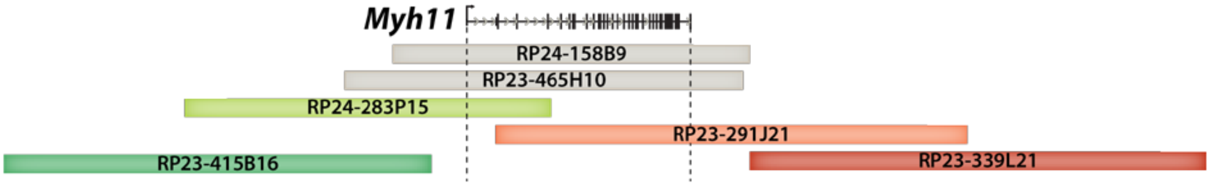
C, Left: Hi-C of all 6 chromosomes shows their territoriality as patches. Second-left: A Hi-C contact map averaged over all 6 chromosomes exhibits the checkerboard pattern typical segregation of A- and B-type chromatin. The three rightmost graphs show zoom views of our genes with stiffness profiles above the maps**.**

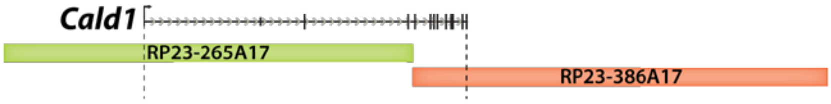
Supplementary Information Table S3. BAC clones used for DNA- and RNA-FISH in this study

<table>
<tr><th colspan="2"><i>Tg</i></th></tr>
<tr><td>Gene body</td><td>RP24-229C15</td></tr>
<tr><td>Gene halves</td><td>RP23-193A18, RP23-266I10, RP24-179P7, RP23-255H1</td></tr>
<tr><td>Gene flanks</td><td>RP24-171I15, RP24-312F21</td></tr>
<tr><td colspan="2">
Tg
RP24-229C15
RP23-193A18
RP23-255H1
RP23-266I10
RP24-179P7
RP24-171I15
RP24-312F21

</td></tr>
<tr><th colspan="2"><i>Ttn</i></th></tr>
<tr><td>Gene body</td><td>RP23-310F9, RP23-22E10</td></tr>
<tr><td>Gene halves</td><td>RP23-140P12, RP23-314B24</td></tr>
<tr><td>Gene flanks</td><td>RP23-1M1, RP23-316G15</td></tr>
<tr><td colspan="2">
Ttn
RP23-310F9
RP23-22E10
RP23-140P12
RP23-314B24
RP23-1M1
RP23-316G15

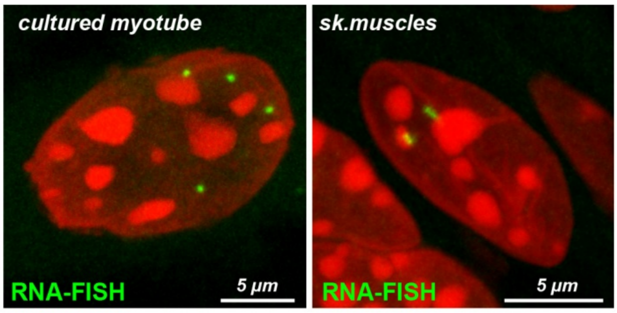
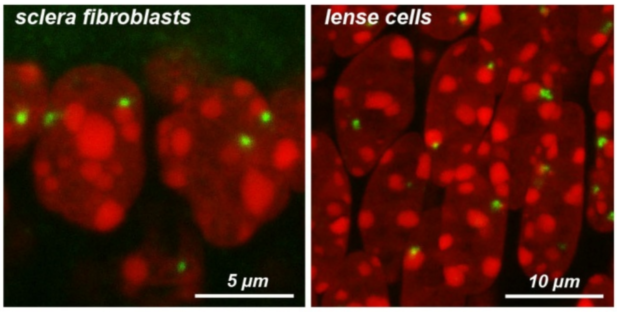
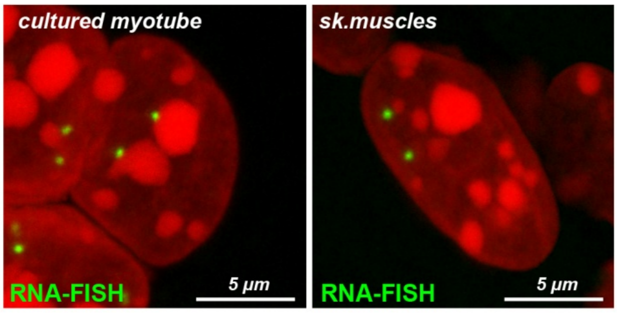
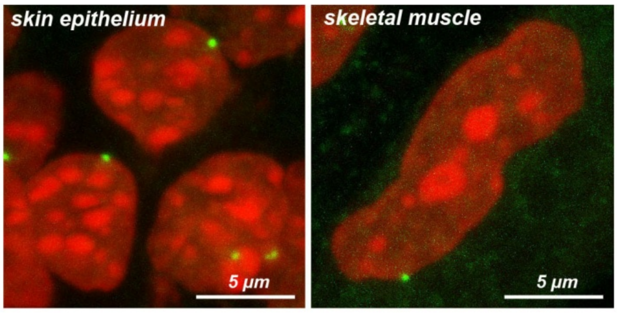
</td></tr>
<tr><th colspan="2"><i>Neb</i></th></tr>
<tr><td>Gene body</td><td>RP23-82J15</td></tr>
<tr><td>Gene halves</td><td>RP23-91F4, RP23-177B19</td></tr>
<tr><td colspan="2">
Neb
RP23-82J15
RP23-91F4
RP23-177B19

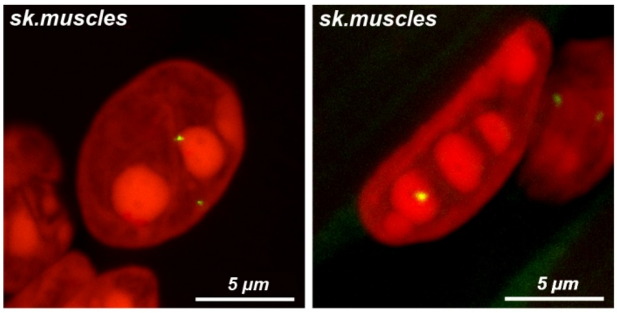
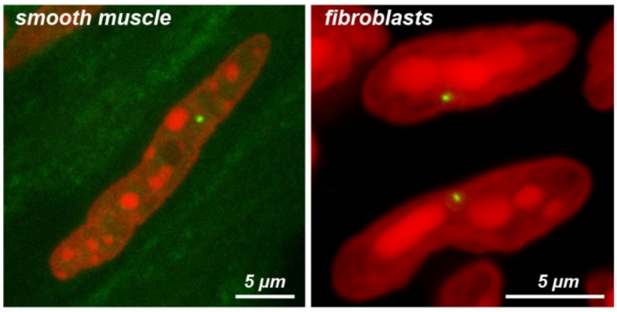
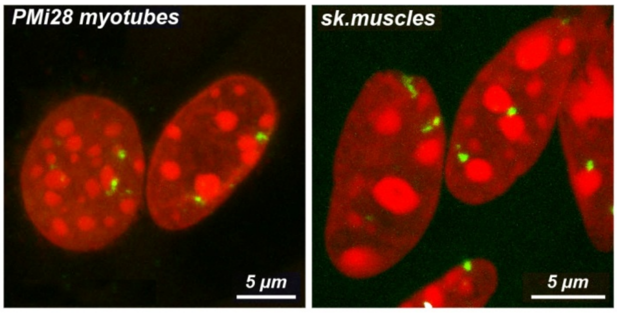
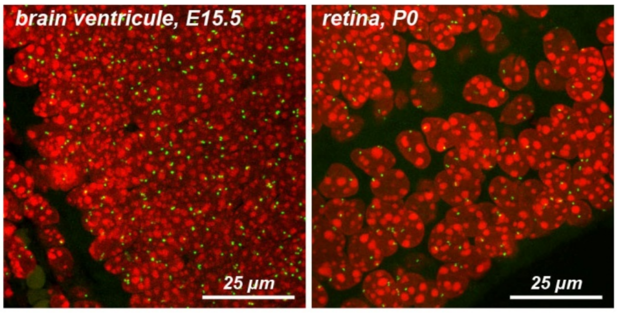
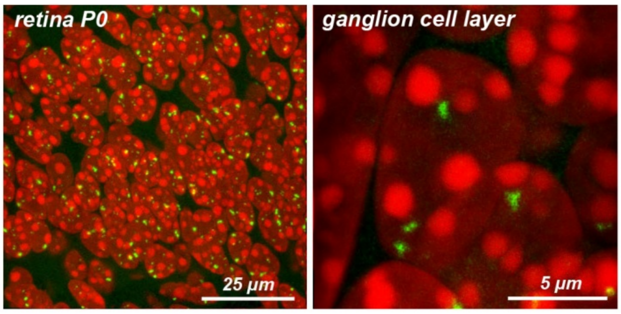
</td></tr>
<tr><th colspan="2"><i>Myh11</i></th></tr>
<tr><td>Gene body</td><td>RP24-158B9, RP23-465H10</td></tr>
<tr><td>Gene halves</td><td>RP24-283P15, RP23-291J21</td></tr>
<tr><td>Gene flanks</td><td>RP23-415B16, RP23-339L21</td></tr>
<tr><td colspan="2">
Myh11
RP24-158B9
RP23-465H10
RP24-283P15
RP23-291J21
RP23-415B16
RP23-339L21

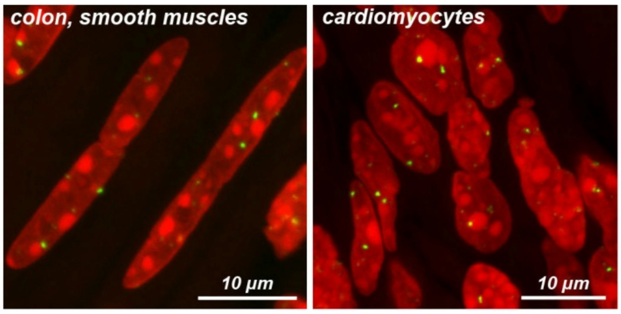
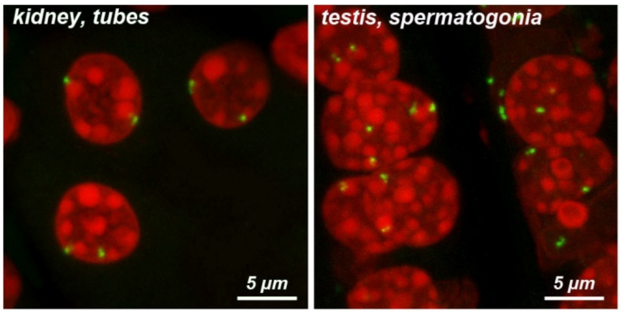
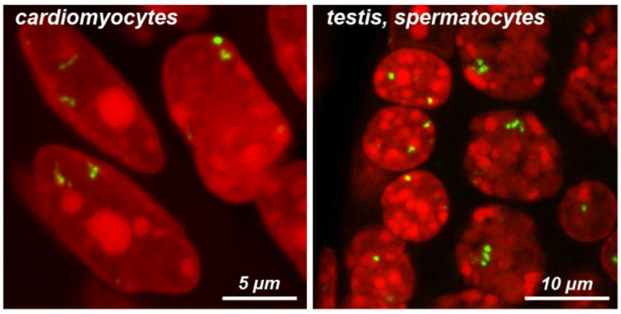
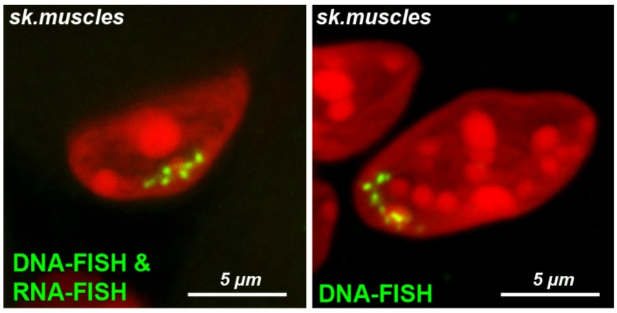
</td></tr>
<tr><th colspan="2"><i>Cald1</i></th></tr>
<tr><td>Gene body</td><td>RP23-265A17 + RP23-386A17</td></tr>
<tr><td colspan="2">
Cald1
RP23-265A17
RP23-386A17

</td></tr>
</table>

Supplementary Information Table S4A. Long lowly or moderately expressed genes do not form TLs resolved by microscopy

gene chromosome length / expression* BAC probe(s)	FISH** in various mouse tissues***: nuclei stained with DAPI (red), BAC signals (green)
Myom1 MMU 17 107 kb RP23-460N24	 cultured myotube sk.muscles RNA-FISH 5 µm RNA-FISH 5 µm
Col4a1 & Col4a2 MMU 8 114 & 137 kb RP23-110G10	 sclera fibroblasts lense cells 5 µm 10 µm
Myo1b MMU 1 214 kb / 19 TPM RP23-395I14	 cultured myotube sk.muscles RNA-FISH 5 µm RNA-FISH 5 µm
Col4a5 MMU X 214 kb RP23-83N2	 skin epithelium skeletal muscle 5 µm 5 µm

Syne2 **MMU 12** **293 kb** **RP24-262E7** **RP24-70E14**	 sk.muscles 5 µm sk.muscles 5 µm
Col4a6 **MMU X** **309 kb** **RP24-93G22**	 smooth muscle 5 µm fibroblasts 5 µm
Syne1 **MMU 10** **535 kb** **RP23-83G17** **RP23-479M12** **RP24-366J3**	 PMi28 myotubes 5 µm sk.muscles 5 µm
Gpr98 **MMU 13** **538 kb** **RP23-110O5** **RP23-87H18** **RP23-112F16**	 brain ventricule, E15.5 25 µm retina, P0 25 µm
Nrxn1 **MMU 17** **1,056 kb** **RP23-137C8** **RP23-471F15** **RP24-82E21** **RP23-119G24** **RP23-141M11** **RP24-273L1**	 retina P0 25 µm ganglion cell layer 5 µm

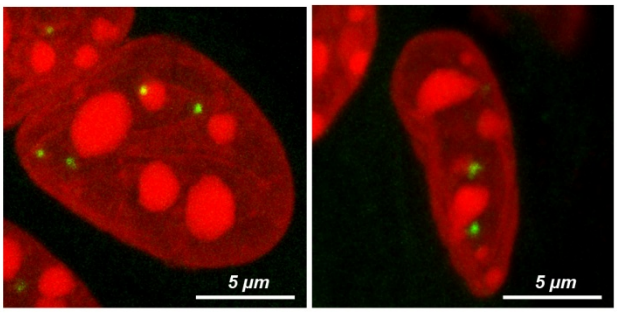
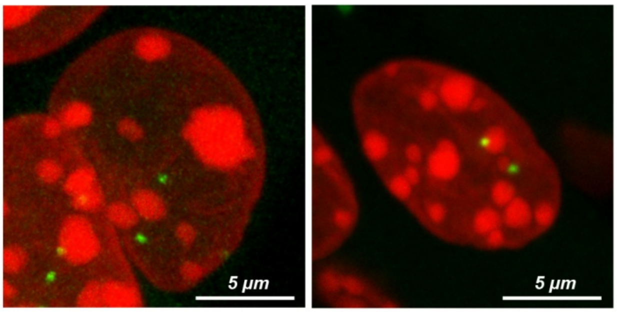
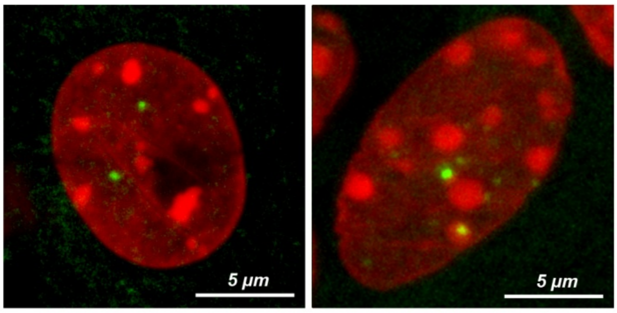
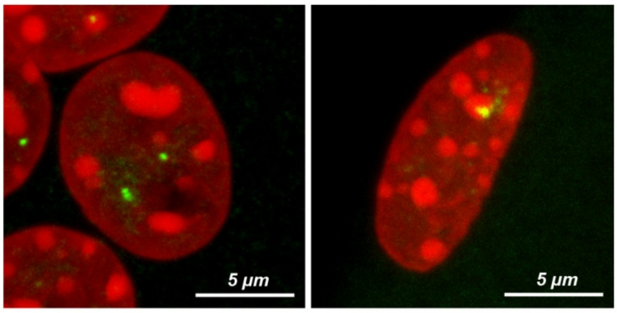
Prkg1 **MMU 19** **1,196 kb** **RP23-429B21** **RP23-13B19** **RP23-266H12** **RP24-375O7**	 colon, smooth muscles 10 µm cardiomyocytes 10 µm
Gpc5 **MMU 14** **1,433 kb** **RP23-392P5** **RP23-336K1** **RP23-26D20** **RP23-1D16**	 kidney, tubes 5 µm testis, spermatogonia 5 µm
Ctnna3 **MMU 10** **1,499 kb** **RP23-116J15** **RP23-25M20** **RP23-121P19** **RP23-155C12** **RP24-209O2**	 cardiomyocytes 5 µm testis, spermatocytes 10 µm
Dmd **MMU X** **2,257,270 kb / 6 TPM** **RP23-39F15** **RP23-35L6** **RP23-191F11** **RP23-386E22** **RP23-121C23** **RP24-70A8**	 sk.muscles DNA-FISH & RNA-FISH 5 µm sk.muscles DNA-FISH 5 µm

* expression level is indicated only for genes expressed in mouse cultured myotubes assessed in this study

** If not indicated otherwise, FISH was carried at conditions when both RNA and DNA are detected

*** selection of tissues for FISH was based on The Human Protein Atlas (https://www.proteinatlas.org/ENSG00000075624-ACTB/tissue#top) and Human Gene DatabaseGeneCards (https://www.genecards.org/)

Supplementary Information Table S4B. Short highly expressed genes do not form TLs resolved by microscopy

gene chromosome length / expression* BAC probe(s)	**RNA-FISH in mouse myotubes (left) & skeletal muscle (right): nuclei stained with DAPI (red), BAC signals (green)**
Acta1 **MMU 8** **3 kb / 4,360 TPM** **RP23-427K5**	
Tpm2 **MMU 4** **9 kb / 2,780 TPM** **RP24-253N9**	
Mb **MMU 15** **35 kb / 550 TPM** **RP23-424C13**	
Myl4 **MMU 11** **45 kb / 6,200 TPM** **RP24-253N9**	

* expression estimated in cultured myotubes

REFERENCES

Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics *36*, 311-316.

Ansari, A., and Hampsey, M. (2005). A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. Genes Dev *19*, 2969-2978.

Anton, T., Bultmann, S., Leonhardt, H., and Markaki, Y. (2014). Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. Nucleus *5*, 163-172.

Bagnoli, J.W., Ziegenhain, C., Janjic, A., Wange, L.E., Vieth, B., Parekh, S., Geuder, J., Hellmann, I., and Enard, W. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat Commun *9*, 2937.

Becker, P.B., and Workman, J.L. (2013). Nucleosome remodeling and epigenetics. Cold Spring Harb Perspect Biol *5*.

Belaghzal, H., Dekker, J., and Gibcus, J.H. (2017). Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. Methods *123*, 56-65.

Bensaude, O. (2011). Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? Transcription *2*, 103-108.

Bienko, M., Crosetto, N., Teytelman, L., Klemm, S., Itzkovitz, S., and van Oudenaarden, A. (2013). A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. Nat Methods *10*, 122-124.

Bjork, P., and Wieslander, L. (2015). The Balbiani Ring Story: Synthesis, Assembly, Processing, and Transport of Specific Messenger RNA-Protein Complexes. Annu Rev Biochem *84*, 65-92.

Branco, M.R., and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol *4*, e138.

Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A., Smith, A.J., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R.*, et al.* (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. J Cell Biol *182*, 1083-1097.

Bystricky, K., Heun, P., Gehlen, L., Langowski, J., and Gasser, S.M. (2004). Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. Proc Natl Acad Sci U S A *101*, 16495-16500.

Carmo-Fonseca, M., and Kirchhausen, T. (2014). The timing of pre-mRNA splicing visualized in real-time. Nucleus *5*, 11-14.

Chakalova, L., and Fraser, P. (2010). Organization of transcription. Cold Spring Harb Perspect Biol *2*, a000729.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., E, P.R.I., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J.*, et al.* (2015). Highly efficient Cas9-mediated transcriptional programming. Nat Methods *12*, 326-328.

Chubb, J.R., Trcek, T., Shenoy, S.M., and Singer, R.H. (2006). Transcriptional pulsing of a developmental gene. Curr Biol *16*, 1018-1025.

Cook, P.R. (1999). The organization of replication and transcription. Science *284*, 1790-1795.

Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45-54.

Cremer, M., Grasser, F., Lanctot, C., Muller, S., Neusser, M., Zinner, R., Solovei, I., and Cremer, T. (2008). Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. Methods Mol Biol *463*, 205-239.

Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harb Perspect Biol *2*, a003889.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193-196.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev *113*, 8456-8490.

Farnung, L., Vos, S.M., and Cramer, P. (2018). Structure of transcribing RNA polymerase II-nucleosome complex. Nat Commun *9*, 5432.

Feodorova, Y., Falk, M., Mirny, L.A., and Solovei, I. (2020). Viewing Nuclear Architecture through the Eyes of Nocturnal Mammals. Trends Cell Biol.

Fujita, K., Iwaki, M., and Yanagida, T. (2016). Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. Nat Commun *7*, 13788.

Gaykalova, D.A., Kulaeva, O.I., Volokh, O., Shaytan, A.K., Hsieh, F.K., Kirpichnikov, M.P., Sokolova, O.S., and Studitsky, V.M. (2015). Structural analysis of nucleosomal barrier to transcription. Proc Natl Acad Sci U S A *112*, E5787-5795.

Grond, C.J., Siegmund, I., and Hennig, W. (1983). Visualization of a lampbrush loop-forming fertility gene in Drosophila hydei. Chromosoma *88*, 50-56.

Hajjoul, H., Mathon, J., Ranchon, H., Goiffon, I., Mozziconacci, J., Albert, B., Carrivain, P., Victor, J.M., Gadal, O., Bystricky, K., *et al.* (2013). High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. Genome Res *23*, 1829-1838.

Hampsey, M., Singh, B.N., Ansari, A., Laine, J.P., and Krishnamurthy, S. (2011). Control of eukaryotic gene expression: gene loops and transcriptional memory. Adv Enzyme Regul *51*, 118-125.

Herzel, L., Straube, K., and Neugebauer, K.M. (2018). Long-read sequencing of nascent RNA reveals coupling among RNA processing events. Genome Res *28*, 1008-1019.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13-23.

Hutchison, N. (1987). Lampbrush chromosomes of the chicken, Gallus domesticus. J Cell Biol *105*, 1493-1500.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods *9*, 999-1003.

Jjingo, D., Huda, A., Gundapuneni, M., Marino-Ramirez, L., and Jordan, I.K. (2011). Effect of the transposable element environment of human genes on gene length and expression. Genome Biol Evol *3*, 259-271.

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat Biotechnol *30*, 90-98.

Kaufmann, R., Cremer, C., and Gall, J.G. (2012). Superresolution imaging of transcription units on newt lampbrush chromosomes. Chromosome Res *20*, 1009-1015.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol *3*, 570-575.

Kulaeva, O.I., Hsieh, F.K., and Studitsky, V.M. (2010). RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. Proc Natl Acad Sci U S A *107*, 11325-11330.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, A., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. Nature.

Lee, K., Hsiung, C.C., Huang, P., Raj, A., and Blobel, G.A. (2015). Dynamic enhancer-gene body contacts during transcription elongation. Genes Dev *29*, 1992-1997.

Li, J., Dong, A., Saydaminova, K., Chang, H., Wang, G., Ochiai, H., Yamamoto, T., and Pertsinidis, A. (2019). Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. Cell *178*, 491-506 e428.

Link, S., Spitzer, R.M.M., Sana, M., Torrado, M., Volker-Albert, M.C., Keilhauer, E.C., Burgold, T., Punzeler, S., Low, J.K.K., Lindstrom, I.*, et al.* (2018). PWWP2A binds distinct chromatin moieties and interacts with an MTA1-specific core NuRD complex. Nat Commun *9*, 4300.

Liu, X., Bushnell, D.A., and Kornberg, R.D. (2013). RNA polymerase II transcription: structure and mechanism. Biochim Biophys Acta *1829*, 2-8.

Liu, X., Farnung, L., Wigge, C., and Cramer, P. (2018). Cryo-EM structure of a mammalian RNA polymerase II elongation complex inhibited by alpha-amanitin. J Biol Chem *293*, 7189-7194.

Liu, Y., Zhou, K., Zhang, N., Wei, H., Tan, Y.Z., Zhang, Z., Carragher, B., Potter, C.S., D'Arcy, S., and Luger, K. (2020). FACT caught in the act of manipulating the nucleosome. Nature *577*, 426-431.

Lucas, J.S., Zhang, Y., Dudko, O.K., and Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. Cell *158*, 339-352.

Macgregor, H.C. (1993). An introduction to animal cytogenetics (Chapman & Hall).

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Martin, M., Cho, J., Cesare, A.J., Griffith, J.D., and Attardi, G. (2005). Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis. Cell *123*, 1227-1240.

Miller, O.L., and Beatty, B.R. (1969). Visualization of nuclear genes. Science *164*, 955-957.
Morgan, G.T. (2018). Imaging the dynamics of transcription loops in living chromosomes. Chromosoma.

Muller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nat Rev Genet *14*, 275-287.

Nicolas, D., Phillips, N.E., and Naef, F. (2017). What shapes eukaryotic transcriptional bursting? Mol Biosyst *13*, 1280-1290.

Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M. (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. Mol Cell *72*, 369-379 e364.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L.A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc Natl Acad Sci U S A *115*, E6697-E6706.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Olins, D.E., Olins, A.L., Levy, H.A., Durfee, R.C., Margle, S.M., Tinnel, E.P., and Dover, S.D. (1983). Electron microscope tomography: transcription in three dimensions. Science *220*, 498-500.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., *et al.* (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet *36*, 1065-1071.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science *357*.

Papantonis, A., and Cook, P.R. (2013). Transcription factories: genome organization and gene regulation. Chem Rev *113*, 8683-8705.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. Gigascience *7*.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., *et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell *38*, 603-613.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Punzeler, S., Link, S., Wagner, G., Keilhauer, E.C., Kronbeck, N., Spitzer, R.M., Leidescher, S., Markaki, Y., Mentele, E., Regnard, C., *et al.* (2017). Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation. EMBO J *36*, 2263-2279.

Rau, A., Gallopin, M., Celeux, G., and Jaffrezic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics *29*, 2146-2152.
Rippe, K. (2001). Making contacts on a nucleic acid polymer. Trends Biochem Sci *26*, 733-740.

Savova, V., Chun, S., Sohail, M., McCole, R.B., Witwicki, R., Gai, L., Lenz, T.L., Wu, C.T., Sunyaev, S.R., and Gimelbrant, A.A. (2016). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. Nat Genet *48*, 231-237.

Savova, V., Vigneau, S., and Gimelbrant, A.A. (2013). Autosomal monoallelic expression: genetics of epigenetic diversity? Curr Opin Genet Dev *23*, 642-648.

Schmidt, C.S., Bultmann, S., Meilinger, D., Zacher, B., Tresch, A., Maier, K.C., Peter, C., Martin, D.E., Leonhardt, H., and Spada, F. (2012). Global DNA hypomethylation prevents consolidation of differentiation programs and allows reversion to the embryonic stem cell state. PLoS One *7*, e52629.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet *20*, 437-455.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., *et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet *42*, 53-61.

Singh, B.N., and Hampsey, M. (2007). A transcription-independent role for TFIIB in gene looping. Mol Cell *27*, 806-816.

Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nat Struct Mol Biol *16*, 1128-1133.

Solovei, I. (2010). Fluorescence in situ hybridization (FISH) on tissue cryosections. In Methods Mol Biol, pp. 71-82.

Solovei, I., and Cremer, M. (2010). 3D-FISH on cultured cells combined with immunostaining. In Methods Mol Biol, pp. 117-126.

Solovei, I., Thanisch, K., and Feodorova, Y. (2016). How to rule the nucleus: divide et impera. Curr Opin Cell Biol *40*, 47-59.

Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.M., Robert, M.C., Capozi, S., Chouaib, R., Backer, V., Mateos-Langerak, J., *et al.* (2016). A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. Nat Commun *7*, 12248.

van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780-791.

Walter, J., Joffe, B., Bolzer, A., Albiez, H., Benedetti, P.A., Muller, S., Speicher, M.R., Cremer, T., Cremer, M., and Solovei, I. (2006). Towards many colors in FISH on 3D-preserved interphase nuclei. Cytogenet Genome Res *114*, 367-378.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zheng, M., Tian, S.Z., Capurso, D., Kim, M., Maurya, R., Lee, B., Piecuch, E., Gong, L., Zhu, J.J., Li, Z.*, et al.* (2019). Multiplex chromatin interactions with single-molecule precision. Nature.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell *65*, 631-643 e634.

STAR METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-RNA polymerase II CTD repeat YSPTSPS antibody (8WG16)	Abcam	Cat# Ab817 RRID: AB_306327
Rabbit anti-Cas9	Clontech	Cat# 632607
Mouse anti-GFP	Roche	Cat# 11814460001 RRID: AB_390913
Donkey anti-mouse Alexa 488	Invitrogen	Cat# 21202 RRID: AB_141607
Donkey anti-mouse Alexa 555	Invitrogen	Cat# 31570 RRID: AB_2536180
Donkey anti-rabbit Alexa 488	Jackson Immuno Research	Cat# 711-547-003 RRID: AB_2340620
Donkey anti-rabbit Alexa 555	Invitrogen	Cat# A-31572 RRID: AB_162543
Streptavidin Alexa 555	Invitrogen	Cat# S32355 RRID: AB_2571525
Sheep anti-Digoxigenin FITC	Roche	Cat# 11207750910 RRID: AB_514501
Biological Samples		
mouse tissues: thyroid and bladder from CD-1 mice	This manuscript	N/A
Chemicals, Peptides, and Recombinant Proteins		
α-amanitin	Sigma-Aldrich	Cat# A2263
Actinomycin D	Sigma-Aldrich	Cat# A9415
5,6-dichloro-1-β-D-ribofuranosylbenzimidazol (DRB)	Sigma-Aldrich	Cat# D1916
Pladienolide B	Cayman Chemical	Cat# 16538
Poly-L-lysine	Sigma-Aldrich	Cat# P1399
Aminoallyl dUTP	Sigma-Aldrich	Cat# A0410
Biotin succinimidyl ester	Invitrogen	Cat# B1582
Digoxigenin succinimidyl ester	Invitrogen	Cat# A2952
FITC succinimidyl ester	Invitrogen	Cat# F6130
Cy3 mono NHS ester	Amersham Biosciences	Cat# PA13101
TAMRA succinimidyl ester	Invitrogen	Cat# T6105
Texas Red succinimidyl ester	Invitrogen	Cat# T6134
Cy5 mono NHS ester	Amersham Biosciences	Cat# PA15101
Mouse Cot-1 DNA	Invitrogen	Cat# 18440016
Salmon sperm DNA	Invitrogen	Cat# 15632011
Paraformaldehyde	Merck	Cat# 1040051000
Vectashield mounting medium	Vector Laboratories	Cat# H-1000-10
Nutrient Mixture F-10 Ham	Sigma-Aldrich	Cat# N6908
Fetal bovine serum (FBS)	Biochrom	Cat# S0115
Horse serum	Thermo Fisher Scientific	Cat# 26050070

Penicillin/Streptomycin	Sigma-Aldrich	Cat# P4333
Maxima H Minus reverse transcriptase and buffer	Thermo Fisher Scientific	Cat# EP0753
Exonuclease I and buffer	New England Biolabs	Cat# M0293S
DNase/RNase-Free Distilled Water	Thermo Fisher Scientific	Cat# 10977-049
Micrococcal nuclease (MNase)	Sigma-Aldrich	Cat# N3755
cOmplete Protease Inhibitor (PI) Cocktail	Roche	Cat# 11697498001
DTT	Roth	Cat# 6908.1
Proteinase K	Invitrogen	Cat# AM2546
Glycogen	Thermo Fisher Scientific	Cat# R0561
Roti-Phenol/Chloroform/Isoamylalcohol	Roth	Cat# A156.1
Halt protease inhibitor cocktail	Thermo Fisher Scientific	Cat# 78438
DpnII	New England Biolabs	Cat# R0543M
ClaI	New England Biolabs	Cat# R0197L
Biotin-14-dATP	Invitrogen	Cat# 19524016
Klenow polymerase	New England Biolabs	Cat# M0210L
Klenow polymerase (exo-)	New England Biolabs	Cat# M0212L
NEBuffer 3.1	New England Biolabs	Cat# B7203S
NEBuffer 2.1	New England Biolabs	Cat# B7202S
CutSmart Buffer	New England Biolabs	Cat# B7204S
T4 DNA ligase buffer	Invitrogen	Cat# 46300018
T4 DNA ligase	Invitrogen	Cat# 15224090
T4 DNA polymerase	New England Biolabs	Cat# M0203L
T4 polynucleotide kinase	New England Biolabs	Cat# M0201L
Critical Commercial Assays		
NucleoSpin Gel and PCR clean-up Kit	Macherey-Nagel	Cat# 740609.250
NucleoSpin RNA Kit	Macherey-Nagel	Cat# 740955.50
NucleoBond Xtra Midi Kit	Macherey-Nagel	Cat# 740410.50
DNeasy Blood & Tissue Kit	Qiagen	Cat# 69504
Qubit dsDNA HS Kit	Invitrogen	Cat# Q32854
MinElute Gel Extraction Kit	Qiagen	Cat# 28606
MaXtract High Density tubes	Qiagen	Cat# 129056
Bioanalyzer High-sensitivity DNA Kit	Agilent Technologies	Cat# 5067-4626
Bioanalyzer DNA 1000 Kit	Agilent Technologies	Cat# 5067-1504
TG Nextera XT DNA Sample Preparation Kit	Illumina	Cat# FC-131-1096
TruSeq Nano DNA Low Throughput Library Prep Kit	Illumina	Cat# 20015964
Dynabeads Protein G	Invitrogen	Cat# 10004D

Dynabeads MyOne Streptavidin C1	Invitrogen	Cat# 65002
CleanPCR SPRI beads	CleanNA	Cat# CPCR-0005
AMPure XP beads	Beckman Coulter	Cat# A63880
NEBNext Multiplex Oligos for Illumina	New England Biolabs	Cat# E7335S
NEBNext Ultra II DNA Library Prep Kit for Illumina	New England Biolabs	Cat# E7645S
DNA Clean & Concentrator-5 Kit	Zymo Research	Cat# D4004
KAPA HiFi HotStart ReadyMix	KAPA Biosystems	Cat# KR0370
Quant-iT PicoGreen dsDNA Assay Kit	Thermo Fisher Scientific	Cat# P11496
High capacity cDNA reverse transcription Kit	Applied Biosystems	Cat# 4368814
LightCycler 480 SYBR Green I Master	Roche	Cat# 04707516001
KAPA SYBR FAST	KAPA Biosystems	Cat# KK4602
Ulysis Nucleic Acid Labeling Kit	Thermo Fisher Scientific	Cat# U21650, Cat# U21652, Cat# U21654, Cat# U21660
GenomiPhi DNA amplification Kit	GE Healthcare	Cat# 25-6600-31
Amicon Ultra column 30K	Millipore	Cat# UFC9 030 0
Deposited Data		
ChIP-seq data	This manuscript	ArrayExpress: E-MTAB-9060
RNA-seq data	This manuscript	ArrayExpress: E-MTAB-9060
Hi-C data	This manuscript	GEO: GSE150704
Experimental Models: Cell Lines		
M. musculus: Pmi28 myoblasts wt	M. C. Cardoso (Technische Universität Darmstadt)	N/A
M. musculus: Pmi28 myoblasts stably expressing dCas9-VPR	This manuscript	N/A
Experimental Models: Organisms/Strains		
M. musculus: CD-1 for organ removal	Charles River Laboratories	Strain code 022
Oligonucleotides		
qPCR primer: Gapdh fwd 5'- CATGGCCTTCCGTGTTCCTA -3'	(Schmidt et al., 2012)	N/A
qPCR primer: Gapdh rev 5'- CTTCACCACCTTCTTGATGTCATC -3'	(Schmidt et al., 2012)	N/A
qPCR primer: Ttn fwd 5'- GCCGCGCTAGATTGATGATC -3'	This manuscript	N/A
qPCR primer: Ttn rev 5'- TCTCGGCTGTCACAAGAAGCT -3'	This manuscript	N/A
qPCR primer: Neb fwd 5'- CAGCAGATGCAGAGTGGGAA -3'	This manuscript	N/A

qPCR primer: Neb rev 5’- TAACAATGCTGGCGTGACCT -3’	This manuscript	N/A
Primer to amplifiy *Tg* exon FISH probe: Tg Ex 2–12 fwd 5’- AGGTAGATGCACAGCCACTC -3’	This manuscript	N/A
Primer to amplifiy *Tg* exon FISH probe: Tg Ex 2–12 rev 5’- CAGGTTCCCAGCCTCCAATC -3’	This manuscript	N/A
Primer to amplifiy *Tg* exon FISH probe: Tg Ex 33–47 fwd 5’- TGCTGCACTGGCTTTGGTTT -3’	This manuscript	N/A
Primer to amplifiy *Tg* exon FISH probe: Tg Ex 33–47 rev 5’- AGGAGCAGTCAGCTTGTTTGA -3’	This manuscript	N/A
Primers to introduce protospacer sequences for dCas9-VPR promoter targeting	This manuscript	See Table S5
Primers to amplify FISH probes for specific regions of *Tg* and *Ttn*	This manuscript via http://www.hdfish.eu/	See Table S5
Recombinant DNA		
BAC clones	BACPAC Resources, Oakland children’s hospital	See Tables S3 and S4
PB-TRE-dCas9-VPR	(Chavez et al., 2015)	Addgene plasmid #63800; RRID: Addgene_63800
U6-Ttn-1-GFP-H2A	This manuscript	N/A
U6-Ttn-2-GFP-H2A	This manuscript	N/A
U6-Ttn-3-GFP-H2A	This manuscript	N/A
U6-Ttn-4-GFP-H2A	This manuscript	N/A
U6-Ttn-5-GFP-H2A	This manuscript	N/A
U6-Ttn-6-GFP-H2A	This manuscript	N/A
U6-Neb-1-GFP-H2A	This manuscript	N/A
U6-Neb-3-GFP-H2A	This manuscript	N/A
U6-Neb-4-GFP-H2A	This manuscript	N/A
U6-Neb-5-GFP-H2A	This manuscript	N/A
U6-Neb-6-GFP-H2A	This manuscript	N/A
U6-Neb-7-GFP-H2A	This manuscript	N/A
Tg cDNA clone	transOMIC technologies	BC111467
Software and Algorithms		
Image J v1.43n	NIH	https://imagej.nih.gov/ij/
ImageJ plugins for work with confocal stacks	(Walter et al., 2006)	N/A
R v3.5.3	RDevelopment Core Team	https://www.R-project.org/
RStudio v1.2.1335	RStudio, Inc	https://rstudio.com/
Flanks Measurements	David Hörl	https://github.com/hoerldavid/fish_analysis
zUMIs pipeline	(Parekh et al., 2018)	N/A
HTS filter	(Rau et al., 2013)	N/A

Bowtie v1.2.2	(Langmead and Salzberg, 2012)	N/A
MACS callpeak	(Zhang et al., 2008)	N/A
Distiller	Mirny Lab	https://github.com/mirnylab/distiller-nf
Cooler	Mirny Lab	https://github.com/mirnylab/cooler
Cooltools	Mirny Lab	https://github.com/mirnylab/cooltools
Pairtools	Mirny Lab	https://github.com/mirnylab/pairtools
Polymer simulations	Mirny Lab	https://github.com/mirnylab/openmm-polymer-legacy
Other		
GTEx RNA-seq data (median TPM values)	https://gtexportal.org/home/datasets	Version 7
GENCODE gene annotations human	https://www.gencodegenes.org/	Version v19
GENCODE gene annotations mouse	https://www.gencodegenes.org/	Version vM11

1. Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Irina Solovei (irina.solovei@lrz.uni-muenchen.de).

Materials Availability

Plasmids generated in this study are available upon request.

Data and Code Availability

ChIP-seq data and RNA-seq are available at ArrayExpress (EMBL-EBI) under accession E-MTAB-9060. Hi-C data have been uploaded to Gene Expression Omnibus (GEO) and are available under accession GSE150704. Accession numbers are also available in the Key Resources Table. Codes used in this study are listed in the Key Resources Table.

2. Experimental model and subject details

Cell line

Pmi 28 mouse myoblasts. Male. Cell line has not been authenticated.

Mouse strain

CD-1. Females and males. Ages ranging from P14 to adult. Purchased from Charles River Laboratories. The mice were exposed to a 12-hour light/12-hour dark cycle and given food and water ad libitum. Mice were housed at the Biocenter, Ludwig Maximilians University of Munich (LMU) and treated according to the standard protocol approved by the Animal Ethics Committee of LMU and The Committee on Animal Health and Care of the local governmental body of the state of Upper Bavaria, Germany.

3. Method details

GTEx data analysis

RNA-seq data on human tissues was downloaded from the website of the GTEx portal (https://gtexportal.org/home/datasets, version 7, median gene level TPMs). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. GENCODE gene annotation version v19 (human) was downloaded from https://www.gencodegenes.org/. Using R (https://www.R-project.org/), mitochondrial genes were excluded from the analysis and only protein-coding genes with a known gene status were selected to ensure proper annotation and used for further evaluation. Gene lengths were calculated from GENCODE annotated start and end positions of each gene. GENCODE annotations were joined with GTEx RNA-seq data based on the genes' Ensembl gene ID.

Cell culture

Mouse myoblast cell line Pmi28 was grown in Nutrient Mixture F-10 Ham supplemented with 20% fetal bovine serum (FBS) and 1% Penicillin/Streptomycin at 37°C and 5% CO_2. For differentiation, cells were seeded at a density of 8 x 10^4 cells/cm^2 and incubated in a differentiation medium (high glucose DMEM supplemented with 1% Penicillin/Streptomycin and 2% horse serum) at 37°C and 5% CO_2. Differentiation medium was replaced every 2 days. Typically, after 3-4 days a sufficient number of myotubes formed and after 5-6 days they started to contract.

Manipulation of transcription and splicing

Transcription stimulation in myoblasts. Transcription stimulation was performed using a system developed by (Chavez et al., 2015). Pmi28 mouse myoblasts were transfected with PB-TRE-dCas9-VPR (purchased from Addgene, ID 63800) and PiggyBac transposase (System Biosciences, #PB200PA-1) expression plasmids at a ratio of 3:1 with Lipofectamine 3000 (Invitrogen) according to manufacturer's instructions. Clones that stably integrated dCas9-VPR were selected with medium containing 50 µg/ml hygromycin B (Sigma-Aldrich) over a time course of 2 weeks. Single cells were sorted into 96 well plates by fluorescence activated cell sorting (FACS) (BD Biosciences Aria II Sorter) and cultured for one week in presence of hygromycin B. dCas9-VPR expression was induced with doxycyclin at a final concentration of 1 µg/ml for 50 h. Individual clones were tested for dCas9-expression by western blot using a Cas9 antibody (Clontech, dilution 1:1000). To verify that dCas9-VPR could be correctly targeted to a desired genomic region, stable Pmi28 clones were transfected with gRNA expression plasmids containing the protospacer sequences for major and minor satellites and telomere repeats, respectively, as described previously (Anton et al., 2014). Localization of dCas9-VPR was assessed via immunofluorescence staining with Cas9 antibody (Clontech, dilution 1:150).

To monitor gRNA transfection efficiency, a plasmid expressing GFP-H2A additionally to the respective gRNA (U6-gRNA-GFP-H2A) was generated. CMV-GFP-H2A was inserted into the pEX-A-U6-gRNA expression plasmid via Gibson assembly. pEX-A-U6-gRNA was synthesized at Eurofins MWG Operon according to (Mali et al., 2013). Gibson assembly was performed using pEX-A-U6-gRNA linearized by SacI and GFP-H2A generated by PCR using the following primers. Fwd: 5'-atatgggtaccgagctTAGTTATTAATAGTAATCAATTACGGG-3'; Rev: 5'-ttgcggccgcgagctTGATCAGTTATCTAGATCCGG-3'. gRNA vectors were generated by amplifying U6-gRNA-GFP-H2A with forward and reverse primers which introduced the protospacer sequence for the respective promoter target region. Primer sequences for the introduction of protospacer sequences are listed in Supplementary Table S5.

For gRNA transfection, 3x10^5 cells were seeded per 6 well on coverslips and transfected with 2.5 µg of a mix of 6 different gRNA plasmids per target gene (see key resources table) at equimolar amounts with Lipofectamine 3000 (Invitrogen) according to manufacturer's instructions. Medium was removed after 12 h and replaced by medium containing 1 µg/ml doxycyclin to induce dCas9-VPR expression.

Transcription inhibition in myotubes. Differentiation medium was supplemented with 10 µg/ml actinomycin D (Sigma-Aldrich) for 12 or 24 h, or 25 µg/ml α-amanitin (Sigma-Aldrich) for 24 h, or 100 µM 5,6-dichloro-1-β-D-ribofuranosylbenzimidazol (DRB) (Sigma-Aldrich) for 3 or 6 h. Fixation time points of cells under DRB treatment of after DRB removal included 2 min, 25 min, 50 min, 75 min and 3 h.

Splicing inhibition in myoblasts. Cells were incubated with 10 nM or 1 µM of Pladienolide B (Cayman Chemical) for 4 h before fixation. As FISH experiments showed, drug concentration of 1 µM leads to more severe transcription abortion. Therefore the final image analysis was performed after treatment with concentration of 10 nM.

Sample preparation for native ChIP-seq and Hi-C

Pmi28 myoblasts were washed with PBS, detached using 0.25 % Trypsin/1 mM EDTA for 3 min and collected in a centrifuge tube. After centrifugation for 4 min at 1,300 g, cells were washed with PBS and centrifuged again. For native ChIP-seq, the pellet obtained from three 150 mm cell culture plates roughly corresponding to 15 x10^6 myoblasts was directly snap frozen in liquid nitrogen and kept at -80 ºC. For Hi-C, cells were resuspended in medium to a final volume of 10 ml in 15 ml falcon tubes. 666 µl of 16% formaldehyde were added to the tube (resulting in a final concentration of 1% formaldehyde); the tube was briefly shaken and then rotated for exactly 10 min at 21 ºC. To quench the reaction, 593 µl of 2.5 M glycine were added; the tube was briefly shaken and left rotating for 5 min. After quenching, tubes were centrifuged for 5 min at 2,000 g and the supernatant was discarded. Cells were resuspended in 1.5 ml PBS, transferred into 2 ml tubes, centrifuged at 2,000 g for 5 min at RT and the supernatant was discarded. The pellets were snap-frozen in liquid nitrogen and kept at -80 ºC.

Myotube samples were collected as follows in order to separate differentiated myotubes from undifferentiated myoblasts: first, myotubes were gently trypsinized using 0.025 % Trypsin/0.1 mM EDTA; when myotubes detached, leaving the majority of myoblasts still attached to the bottom of the dish, the supernatant containing myotubes and a proportion of detached myoblasts was transferred to a fresh 150 mm cell culture dish. The remaining myoblasts were allowed to re-adhere to the dish bottom, whereas myotubes did not re-adhere and thus could be collected with the supernatant. After centrifugation for 4 min at 1,300 g, cells were washed with PBS and centrifuged again. For native ChIP-seq, the pellet obtained from three 150 mm cell culture plates roughly corresponding to 10 x10^6 myotubes was snap frozen in liquid nitrogen and kept at -80 ºC. For Hi-C, myotubes were fixed, quenched and frozen in the same way as myoblasts. We estimated that our myotube samples contained 10-20% of myoblasts.

Mouse thyroids for Hi-C were prepared as follows: thyroid glands from three mice were isolated within 5-10 min and stored in medium containing TSH (DMEM/F-12; 20% FCS; Penicillin/Streptomycin; 2 mM L-Glut; 40 µg/ml vitamin C; 1 U/ml TSH) in a cell culture incubator at 37 ºC. Glands were then processed one by one. The two lobes of each gland were cleaned from attached neighboring tissues (muscles, fat, parathyroid, etc) under a binocular and placed back in the incubator in the same medium while the other glands were being processed. Finally, all six lobes were joined in 2 ml of DMEM without FCS, minced with micro-scissors into small pieces (5-6 from each lobe) and transferred to a 2 ml tube using low retention tips. The tube was centrifuged at 2,000 rpm for 5 min at RT and the supernatant was removed. Tissue was resuspended in 1.5 ml of DMEM without FCS. 100 µl of 16% formaldehyde were added to the tube (resulting in a final concentration of 1% formaldehyde); the tube was briefly shaken and then rotated for exactly 10 min at 21 ºC. To quench the reaction, 89 µl of 2.5 M glycine were added; the tube was briefly shaken and left rotating for 5 min. Tube was moved to ice and kept there for 0.5 h. After quenching, tubes were centrifuged for 10 min at 6,000 rpm and supernatant was removed. The pellets were snap-frozen in liquid nitrogen and kept at -80 ºC. The same procedure was repeated with three other mice so that each biological replicate consisted of two tubes each containing cells from 6 thyroid lobes. We estimated that thyroid samples contained ca. 60% of thyrocytes.

For smooth muscles, we used a mouse bladder tissue; a bladder per sample. The tissue was minced and fixed as described for thyroid. We estimated that bladder samples contained 40-50% of smooth muscles.

RNA-seq

Total RNA from cells or tissues was isolated using the NucleoSpin RNA Kit (Macherey-Nagel) according to the manufacturer's instructions. To enable proper lysis, tissues were disrupted directly in 350 µl lysis buffer RA1 using a Tissue Homogenizer (Bullet Blender 24, Next Advance) with zirconium oxide beads with a diameter of 0.5 mm for 5 to 10 minutes until a homogeneous solution had formed, prior to applying the sample to the clearing column.

Digital gene expression libraries for RNA-seq were produced using a modified version of single-cell RNA barcoding sequencing (SCRB-seq) optimized to accommodate bulk cells (Bagnoli et al., 2018; Ziegenhain et al., 2017). For each sample, bulk cell input RNA (70 ng) was reverse transcribed in 10 µl reactions containing 25 units of Maxima H Minus reverse transcriptase (Thermo Fisher Scientific), 1x Maxima RT Buffer (Thermo Fisher Scientific), 1 mM dNTPs (Thermo Fisher Scientific), 1 µM E3V6NEXT oligo-dT primer with a sample-specific barcode (Integrated DNA Technologies), and 1 µM E5V6NEXT template-switching oligo (Integrated DNA Technologies). Reverse transcription reactions were incubated 90 min at 42 °C. Next, the cDNA from individual samples was pooled and purified using the DNA Clean & Concentrator-5 Kit (Zymo Research) according to the manufacturer's instructions. Purified cDNA was eluted in 18 µl UltraPure DNase/RNase-Free Distilled Water (Thermo Fisher Scientific). Residual primers were then digested by adding 1 µl Exonuclease I Buffer (New England Biolabs) and 1 µl Exonuclease I (New England Biolabs) (final reaction volume: 20 µl) and incubating at 37 °C for 30 min followed by heat-inactivation at 80 °C for 20 min. Full-length cDNA was then pre-amplified by adding 30 µl PCR master mix consisting of 25 µl KAPA HiFi HotStart ReadyMix (KAPA Biosystems) and 0.33 µM SINGV6 primer (Integrated DNA Technologies). After an initial denaturation at 98 °C for 3 min, the pre-amplification PCR was performed using the following conditions for 10 cycles: 15 s at 98 °C, 30 s at 65 °C, 6 min at 68 °C. Final elongation was performed for 10 min at 72 °C. The pre-amplified cDNA pool was then purified using a 1:0.8 ratio of CleanPCR SPRI beads (CleanNA) with a final elution in 12 µl UltraPure DNase/RNase-Free Distilled Water (Thermo Fisher Scientific).

The concentration of the purified pre-amplified cDNA was then quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific) while the size distribution and quality of the pre-amplified cDNA was assessed with a Bioanalyzer (Agilent Technologies) using a High-sensitivity DNA Kit (Agilent Technologies). Next, a Nextera XT DNA library was constructed with the TG Nextera XT DNA Sample Preparation Kit using 0.8 ng of preamplified cDNA in a final reaction volume of 20 µl containing 4 µl UltraPure DNase/RNase-Free Distilled Water (Thermo Fisher Scientific), 10 µl Tagment DNA Buffer (Illumina) and 10 µl Amplicon Tagment Mix (Illumina). The tagmentation reaction was incubated at 55 °C for 10 min, after which, 5 µl NT buffer (Illumina) was added and the sample was incubated at room temperature for 5 min.

Following tagmentation, a custom P5 primer (P5NEXTPT5) (Integrated DNA Technologies) was used to enrich for 3' cDNA ends in the final Nextera XT Indexing PCR. To this end, the 25 µl tagmentation sample containing cDNA was supplemented with 25 µl of PCR master mix containing 0.1 µM P5NEXTPT5 primer (Integrated DNA Technologies), 0.1 µM Nextera i7 index primer (Integrated DNA Technologies), and 15 µl NPM PCR Mix (Illumina). The Nextera Index PCR was performed using the following conditions: elongation at 72 °C for 3 min followed by an initial denaturation at 95 °C for 30 sec, then 13 cycles: 10 s at 95 °C, 30 s at 55 °C, 1 min at 72 °C. Final elongation was performed at 72 °C for 5 min. The amplified Nextera XT Library was first purified using a 1:1 ratio of CleanPCR SPRI beads (CleanNA) with a final elution in 12 µl UltraPure DNase/RNase-Free Distilled Water (Thermo Fisher Scientific).

The purified Nextera XT Library was size-separated using a 2 % E-Gel Agarose EX Gel (Life Technologies) and the 300–800 bp range of the library was excised from the gel and the DNA extracted using the MinElute Gel Extraction Kit (Qiagen) according to manufacturer's recommendations. The final concentration of the size-selected Nextera XT library was then quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific) while the size distribution and quality of Nextera XT library was assessed with a Bioanalyzer (Agilent Technologies) using a High-sensitivity DNA Kit (Agilent Technologies). Paired-end sequencing of the Nextera XT RNA-seq library was performed using a high output flow cell on an Illumina HiSeq 1500. In the first read, 16 bases were sequenced to obtain the sample-specific barcodes, while the 50 bases in the second read provided the sequence of the cDNA fragment. As multiple libraries were sequenced in parallel on the flow cell, an additional 8 base i7 barcode read was performed.

RNA-seq libraries were processed and mapped to the mouse genome (mm10) using the zUMIs pipeline (Parekh et al., 2018). UMI count tables were filtered for low counts using HTSFilter (Rau et al., 2013). GENCODE gene annotation version vM11 (mouse) was downloaded from https://www.gencodegenes.org/. Using R (https://www.R-project.org/), mitochondrial genes were excluded and only protein-coding genes with a known gene status were selected to ensure proper annotation and used for further evaluation. Gene lengths were calculated from GENCODE annotated start and end positions of each gene.

Native ChIP-seq

The following nChIP-seq protocol was adapted from (Punzeler et al., 2017) and (Link et al., 2018). Immunoprecipitations were carried out with S1 mononucleosomes derived from Pmi28 myoblasts and Pmi28 myotubes. nChIP-seq experiments were performed in independent duplicates.

Per immunoprecipitation, $5x10^6$ cells were used. Cells were resuspended in 750 µl PBS + 1 x PI + 0.3 % Triton X-100 and incubated for 10 min at 4 °C while rotating. The cell lysate was centrifuged for 10 min at 3,000 g and 4°C, the pellet was resuspended in 625 µl PBS + PI and centrifuged for 5 min at 3,000 g at 4°C. Supernatant was carefully removed and the pellet was resuspended in 62.5 µl EX100 (10 mM HEPES pH 7.6, 100 mM NaCl, 1.5 mM $MgCl_2$, 0.5 mM EGTA, 10 % (v/v) glycerol, 10 mM β-glycerol phosphate) + 1x PI + 1 mM DTT and transferred into a DNA low binding tube. $CaCl_2$ concentration was adjusted to 2 mM and 0.1875 µl MNase (1 U/µl) were added. MNase digest was carried out for 20 min at 26 °C. The reaction was stopped by addition of EDTA to a final concentration of 10 mM and the suspension was centrifuged at 13,300 rpm for 30 min at 4 °C in a tabletop centrifuge. The supernatant (S1 mononucleosomes) was transferred into a fresh DNA low binding tube and filled up to 500 µl with EX100 + 1x PI + 1 mM DTT. 25 µl (5 %) were saved as input.

5 µg of antibody (anti-RNA polymerase II CTD repeat YSPTSPS, 8WG16, Abcam ab817) were added per IP ($5x10^6$ cells) in 500 µl and incubated with mononucleosomes at 4 °C overnight while rotating. On the next day, magnetic beads (Dynabeads Protein G, Invitrogen) were washed 3 times with EX100 + 1 x PI + 1 mM DTT (500 µl/wash), resuspended in 30 µl EX100 + 1 x PI + 1 mM DTT per IP, added to the antibody-mononucleosome suspension and rotated at 4 °C for 3 h. A beads-only IP served as negative control.

Washing steps were conducted using 500 µl of buffer per washing step and samples were rotated at 4 °C for 5 min. Beads were magnetically separated after each washing step. Beads were washed twice with WB1 (10 mM Tris pH 7.5, 1 mM EDTA, 0.1 % SDS, 0.1 % sodiumdeoxycholate, 1 % Triton X-100), twice with WB2 (10 mM Tris pH 7.5, 1 mM EDTA, 0.1 % SDS, 0.1 % sodiumdeoxycholate, 1 % Triton X-100, 150 mM NaCl), once with TE + 0.2 % Triton-X 100 and once with TE buffer.

Washed beads were resuspended in 100 µl TE buffer. To the input DNA sample, 75 µl TE were added. 3 µl of 10 % SDS and 5 µl of proteinase K were added and the suspensions were incubated 1 h at 65 °C. After brief vortexing, suspensions were magnetically separated and supernatant was transferred into DNA low binding tubes (Eppendorf). Beads were washed with 100 µl TE + 0.5 M NaCl, vortexed briefly, magnetically separated and the supernatant added to the first supernatant. For input DNA, 100 µl TE + 0.5 M NaCl were added. Eluted IP DNA and input DNA were Phenol/Chloroform/Isoamylalcohol extracted using MaXtract high density tubes (Qiagen) and ethanol precipitated according to a standard protocol. After extraction, the IP DNA pellets were resuspended in 12 µl 10 mM Tris-HCl pH

7.5 and the input DNA pellet was resuspended in 20 µl 10 mM Tris-HCl pH 7.5. DNA concentrations were determined via the Qubit dsDNA High Sensitivity Kit (Invitrogen) and DNA size was analyzed on a DNA 1000 BioAnalyzer chip (Agilent Technologies).

Illumina sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) and the NEBNext Multiplex Oligos for Illumina (New England Biolabs) according to the manufacturer's instructions. The number of amplification cycles was scaled according to the amount of input material, amplification was validated by measuring DNA concentrations using the Qubit dsDNA High Sensitivity Kit (Invitrogen) and library quality was determined using a Bioanalyzer (Agilent Technologies) and a DNA 1000 Bioanalyzer chip (Agilent Technologies).

ChIP-seq libraries were sequenced on an Illumina HiSeq 1500 platform using single read 50 nucleotide sequencing. ChIP–seq reads were aligned to the mouse genome (mm10) with Bowtie (v.1.2.2) (Langmead and Salzberg, 2012) with parameters '-a -m 3 -n 3 --best --strata'. Subsequent ChIP–seq analysis was carried out on data of merged replicates. Signal pile up was performed using MACS2 callpeak (Zhang et al., 2008) with the parameters '--extsize 150' for ChIP, '--extsize 220' for hMeDIP, and '--nomodel -B --nolambda' for all samples.

Hi-C

Hi-C was performed as described by (Belaghzal et al., 2017) with modifications. 3 biological replicates were prepared for each of the studied cell types – myoblasts, myotubes, thyrocytes and smooth muscles. Sample sizes were estimated as following: 15 x10^6 of myoblasts, 10 x10^6 of myotube nuclei, 3 x10^6 of thyroid cells (thyroids of 6 mice); number of cells in a bladder was not estimated.

Formaldehyde fixed Hi-C samples were incubated on ice for 15 min in cold lysis buffer [10 mM Tris-HCl (pH=8.0), 10 mM NaCl, 0.2% (v/v) Igepal (NP40)] mixed with Halt protease inhibitor cocktail (Thermo Fisher Scientific) immediately before. For thyrocytes and smooth muscle, 1 ml of lysis buffer with 10 µl of protease inhibitors was used; myoblasts and myotubes were split into 3 or 2 subsamples, respectively, each containing 5 x10^6 cells. For each subsample 1 ml of lysis buffer was used. Myoblast and myotube subsamples were treated as separate samples until the DNA purification step when they were merged into one biological replicate.

Next, samples were lysed with a Dounce homogenizer and pestle A (KIMBLE Kontes) by moving the pestle slowly up and down 30 times. The procedure was repeated 4 times until 120 strokes with the pestle were performed for each cell sample. Cell suspensions were centrifuged for 5 minutes at 2500 x *g* at RT using a table top centrifuge (Eppendorf). Pellets were washed twice with 500 µl ice cold 1x NEBuffer 3.1 (New England Biolabs) and centrifuged for 5 min at 2500 x *g*. After the second wash, each pellet was resuspended in 360 µl 1x NEBuffer 3.1. Chromatin was solubilized by addition of 38 µl 1% SDS per tube, the mixture was resuspended and incubated at 65°C for 10 minutes. Tubes were placed on ice and 43 µl of 10% Triton X-100 was added. Chromatin was subsequently digested by adding 400 Units *DpnII* (New England Biolabs) [8 µl *DpnII* 50 U/µl, 12 µl NEBuffer 3.1 and water up to 52 µl] at 37°C overnight in a thermo-mixer (Eppendorf) with interval mixing (30 sec ON, 4 min OFF, 900 rpm). Digested chromatin solution was incubated for 20 min at 65°C to deactivate the endonuclease enzyme.

The *DpnII* DNA ends were filled in and marked with biotin by adding 60 µl fill-in mix [2 µl water, 6 µl 10x NEBuffer 3.1, 1.5 µl 10mM dCTP, 1.5 µl 10mM dGTP, 1.5 µl 10mM dTTP, 37.5 µl 0.4mM biotin-14-dATP, 10 µl 5U/ µl Klenow polymerase] followed by incubation for 4 hours at 23°C in a thermo-mixer with interval mixing (15 sec ON, 5 min OFF, 900 rpm). Tubes were placed on ice immediately afterwards and 660 µl of ligation mix [238 µl water, 240 µl 5x T4 DNA ligase buffer (Invitrogen), 120 µl 10% Triton-X100, 12 µl 10mg/ml BSA and 50 µl 1U/µl T4 DNA ligase (Invitrogen)] were added to each sample. Blunt end ligation was performed for 4 hours at 16°C in a thermo-mixer with interval mixing (30 sec ON, 3 min OFF, 900 rpm). Reverse crosslinking was performed by addition of 50 µl of 10 mg/ml Proteinase K to each tube. Samples were incubated for 2 hours at 65°C followed by a second addition of 50 µl of Proteinase K and overnight incubation at 65°C.

Each Hi-C sample was split into three 1.5 ml tubes (ca. 400 µl per tube) and mixed with 1x volume of phenol:chloroform (1:1, pH=8). Tubes were vortexed for 1 min and samples were transferred to 2 ml MaXtract High Density tubes (Qiagen). Samples were centrifuged for 5 min at 16 000 x *g* and the aqueous phase was collected in a 15 ml tube. 120 µl of 3M sodium acetate (pH=5.2) and 3 ml ice-cold 100% ethanol were added to each tube and mixed well by inversion. The volume of each sample was then again split in 3 1.5 ml tubes and incubated for 1 hour at -80°C to precipitate. Samples were then centrifuged at 4°C for 30 min at 16 000 x *g* (Eppendorf); the supernatant was discarded and the pellet air-dried at room temperature. Each pellet was dissolved in 450 µl 1x Tris Low EDTA (TLE) buffer [10mM Tris-HCl, 0.1mM EDTA]. At this point all subsamples of myoblasts and myotubes were fused into 2 samples, respectively. Samples in the 1.5 ml tubes were pooled together in 15 ml tubes, one for each biological replicate. Volume in the 15 ml tubes was brought up to 8 ml with 1x TLE and transferred to a 15 ml 30K Amicon Ultra column (Millipore) for concentrating the samples and removing salts. Tubes were centrifuged at RT for 10 min at 3214 x *g* (Eppendorf). The column was washed with 8 ml 1xTLE and centrifuged again. The volume in the column was brought up to 400 µl by adding 1x TLE and transferred to a 0.5 ml Amicon Ultra 30K column. Columns were centrifuged for 10 min at 21 000 x *g*; 80 µl of 1x TLE was added to each column, the column was flipped into a new collection tube and centrifuged again for 2 min. The volume of all samples at this point was 100 µl. Samples were incubated with 1 µl of 10 mg/ml RNaseA for 1 h at 37°C. DNA concentration was quantified using agarose gel electrophoresis.

Hi-C libraries were treated with T4 DNA polymerase to remove biotinylated ends that did not ligate (dangling ends). The reactions were assembled as follows: Hi-C library (up to 8 µg DNA), 6.5 µl 10x NEBuffer 2.1, 1.25 µl 1 mM dATP, 1.25 µl 1 mM dGTP and 6.5 µl T4 DNA polymerase (3000U/ml, New England Biolabs) in a total volume of 50 µl. 4-5 reactions per sample were mixed in a 1.5 ml tube and then split into PCR tubes. Reactions were incubated at 20°C for 4 hours and then at 75°C for 20 min in a Mastercycler pro S (Eppendorf). The DNA was sheared to a size of 100-500 bp using a Covaris M220 instrument (Covaris), with 130 ul Covaris tubes. The settings were as follows: Peak power 50.0, Duty factor 20.0, Cycles/burst 200, Avg. Power 10.0, Treatment 180 sec. Sheared DNA was size selected in two steps: First, to remove fragments larger than 500bp, each sample (brought to 500ul with 1x TLS) was mixed with 350 µl AMPure XP beads (Beckman Coulter), to a AMPure/Sample ratio of 0.7x, and incubated for 10 min at RT on a rotator. The 0.7x AMPure beads were collected on a magnet for 5 min, and the supernatant (containing DNA below 500bp) was transferred to a new tube for the second size selection with a 1.1x AMPure/Sample ratio, to collect only DNA above 100bp. The beads for the 1.1x step beads were prepared in the following way: 500 µl of beads were transferred to tubes and incubated for 5 min on a magnet; supernatant was discarded and 200 µl of beads were added. After addition of 0.7x supernatant, these were then incubated for 10 min on a rotator. Finally, both the 0.7x and the 1.1x beads were washed twice with 70% ethanol, air-dried and eluted in 25 µl 1x TLE buffer. To recover a greater amount of Hi-C ligation products in the average size range of 200-250bp, the 0.7x eluates containing DNA > 500bp were brought to a volume of 130 µl by adding water and re-sonicated with the same settings but for 120 sec. The same size selection steps were repeated on the re-sonicated samples. By using this modification in the protocol, 3-8 µg of DNA were recovered from the samples, as quantified by agarose gel electrophoresis.

DNA end repair was performed by mixing the size selected DNA eluate with 7 µl of 10x Buffer for T4 ligase with 10 mM ATP (New England Biolabs), 0.7 µl 25 mM dNTP mix, 2.5 µl T4 DNA polymerase (3U/ µl, New England Biolabs), 2. µl T4 polynucleotide kinase (10U/µl, New England Biolabs), 0.5 µl Klenow DNA polymerase I (5U/µl, New England Biolabs) and water up to 70 µl. The reaction was incubated at 20°C for 30 min and at 75°C for 20 min to deactivate the enzymes. Tubes were placed on ice before proceeding to biotin pulldown. All subsequent steps were performed in DNA LoBind tubes (Eppendorf) and each step was performed in a fresh tube. 10-15 µl of streptavidin Dynabeads (MyOne Streptavidin C1 Beads, Invitrogen) were washed twice with 400 µl Tween Wash Buffer (TWB) [5 mM Tris-HCl (pH=8.0), 0.5 mM EDTA, 1 M NaCl, 0.05% Tween20] by incubating for 3 min at RT with rotation, reclaiming against a magnetic separation rack for 1 min and removing all supernatant. Next, reclaimed beads were resuspended in 400 µl 2x Binding Buffer (BB) [10 mM Tris-HCl pH8.0, 1 mM EDTA, 2 M NaCl] and combined with 70 µl Hi-C DNA from the previous step and 330 µl 1x TLE. The mixture was incubated at RT for 15 min with rotation. The supernatant was removed and the DNA-bound Streptavidin beads were washed once with 400 µl 1x BB, once with 100 µl 1x TLE buffer and resuspended in 41 µl of 1x TLE buffer. Samples were kept on ice. Next, the A-tailing reaction was done by mixing the DNA on beads with 5 µl 10x NEBuffer 2.1, 1 µl 10 mM dATP, 3 µl Klenow (exo-) (5U/µl, New England Biolabs) and water in a final volume of 50 µl. The reactions were incubated in a thermocycler at 37°C for 30 min followed by incubation at 65°C for 20 min to inactivate Klenow (exo-) polymerase. The reactions were cooled on ice, transferred to 1.5 ml low-bind tubes and beads were reclaimed against the magnet for 1 min. Beads were washed in 400 µl 1x ligase buffer and resuspended in 36.25 µl 1x ligase buffer.

Adapter ligation reaction was set-up as follows: 36.25 µl Hi-C library on beads, 1 µl 10 mM ATP, 6 µl Illumina TruSeq paired end adapters (TruSeq Nano DNA LT kit, Illumina, San Diego, CA), 2.75 µl 5x DNA ligase buffer (Invitrogen), 4 µl T4 DNA ligase (Invitrogen). The reaction was incubated at RT for 2 hours. The beads with bound ligated Hi-C DNA were collected by holding against a magnet, washed twice with 400 µl 1x TWB for 5 min on a rotator, once with 200 µl 1x BB, once with 200 µl 1x NEBuffer 2.1 to remove non-ligated Paired End adapters and resuspended in 20 µl 1x NEBuffer 2.1.

Two 50 µl reactions were performed for each sample, each containing 6 µl Dynabead-bound Hi-C library, 5 µl PCR primer cocktail, 20 µl Enhanced PCR mix and 19 µl water (TruSeq Nano DNA LT kit, Illumina). 5 or 6 PCR cycles were performed for each biological replica and each cell type, based on the results of a PCR cycle titration test. The temperature profile during the PCR amplification was 3 min at 95°C (initial denaturation), followed by 5/6 cycles of 20 sec at 98°C, 30 sec at 60°C, 30 seconds at 72°C and a final 5-min extension at 72°C. The PCR reactions were pooled together for each biological replicate, Streptavidin beads bound to library templates were collected by magnetic separation for 2 min and supernatants containing amplified sequencing library were transferred to new tubes. Hi-C libraries were size selected from the supernatants using Ampure XP beads as follows: 1.1x volumes of the beads were added to Hi-C samples, briefly vortexed, incubated at RT for 10 min and then collected on a magnet rack for 5 min. Supernatant was discarded and beads were washed twice with 200 µl of freshly made 70% ethanol. Air-dried beads were resuspended in 25 µl water, incubated at RT for 10 min with tapping the tubes every 1-2 min and collected on a magnetic rack for 5 min. Supernatants were transferred to fresh tubes, containing the final Hi-C libraries. Agarose gel electrophoresis was performed to quantify the amount of DNA in each Hi-C library. Samples were diluted in 25 µl water at a final concentration of 10 nM DNA and sent for sequencing. Hi-C libraries were sequenced using 50 bp paired-end sequencing on an Illumina HiSeq 4000.

Hi-C data processing

Fastq files were mapped to the mm10 genome using the distiller pipeline (github.com/mirnylab/distiller-nf). Hi-C interaction data was stored in the cooler format (Abdennur and Mirny, 2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. doi: 10.1093/bioinformatics/btz540). We always applied iterative correction to our data in order to remove biases as much as possible (Imakaev et al., 2012).

Cis/trans contact frequencies were calculated for each bin as the number of cis (same chromosome) contacts divided by the number of trans (different chromosome) contacts of that bin. This is repeated for all bins along the genome, resulting in a genomic track of cis/trans contact frequency ratios.

For a comparison plot of the cis/trans ratio across a number of genes we rescaled the x-axis for each gene so that TSS and TTS align. Interpolation was used to obtain the same number of points in each gene. The cis/trans ratios were normalized to regions outside the genes (to be precise: the region from TSS-3' genelength to TSS-0.5' genelength and TTS+0.5' genelength to TTS+3' genelength is used to normalize cis/trans profile to unity).

This is done to highlight local changes in the cis/trans ratio associated with genes while suppressing longer-range variations.

Fluorescence in situ hybridization (FISH)

BAC-derived FISH probes. BAC clones encompassing the desired genomic region were selected using the UCSC genome browser (https://genome.ucsc.edu/) and purchased from BACPAC Resources (Oakland children's hospital) as agar stabs (https://bacpacresources.org/). Bacteria were grown in LB medium containing 12.5 µg/ml chloramphenicol and BACs were purified via standard alkaline lysis (2 ml of bacterial culture) or the NucleoBond Xtra Midi Kit (Macherey-Nagel) (200 ml of bacterial culture) and amplified using the GenomiPhi Kit (GE Healthcare) according to the manufacturer's instructions. BAC probes were labeled either directly with fluorophores (using fluorophore-dUTPs), or with haptens (biotin-dUTP, digoxigenin-dUTP) by Nick translation.

Labeling of dUTPs and preparation of dNTP mix. Labeling of dUTPs was performed as described previously (Cremer et al., 2008) with modifications. 1 mg of Aminoallyl-dUTP (AA-dUTP, Sigma-Aldrich) was dissolved in 96 µl 0.2 M bicarbonate buffer ($NaHCO_3$), yielding a final concentration of 20 mM. Fluorochromes and haptens were diluted in DMSO to a final concentration of 40 mM (for biotin succinimidyl ester (BIO), digoxigenin succinimidyl ester (DIG) or FITC succinimidyl ester), 20 mM (for Cy3 mono NHS ester or Cy5 mono NHS ester) or 10 mM (for TAMRA succinimidyl ester or Texas Red succinimidyl ester). The reaction to conjugate fluorophores/haptens to dUTPs (10 µl 20 mM AA-dUTP, 10 µl 0.2 M bicarbonate buffer, 10 µl of fluorochrome/hapten dilution in DMSO (20 µl for Texas Red) and 10 µl H_2O) was incubated at RT for 3-4 h. 2 µl 2 M glycine (pH 8.0) and 4 µl 1 M Tris-HCl (pH 7.75) were added to stop the reaction and stabilize the nucleotides. Then, the volume was adjusted to 200 µl with ddH_2O, yielding 1 mM solutions of the respective conjugated dUTP. Labeled nucleotides were stored at –20 °C.

A mix of unlabeled dNTPs and labeled dUTPs was prepared in advance for later use in nick translations. First, a 1 mM solution of dATP/dGTP/dCTP was prepared by adding 4 µl of each dNTP (100 mM, Roche) to 388 µl ddH_2O. Second, a 1 mM solution of dTTP was prepared by adding 4 µl of dTTP (100 mM, Roche) to 396 µl ddH_2O. To prepare the dNTP mix (1 mM), 50 µl 1 mM dATP/dGTP/dCTP, 25 µl 1 mM dTTP and 25 µl 1 mM fluorophore/hapten-labeled dUTP were combined.

Nick translation. BAC DNA was labeled with fluorophores or haptens by nick translation (NT). Typically, 1 µg of DNA (or 4 µl of GenomiPhi amplified BAC DNA) was labeled in a reaction volume of 50 µl (1 µg probe DNA; 5 µl 10x DNA Pol I buffer, Thermo Scientific; 5 µl 0.1 M β-mercaptoetanol; 5 µl dNTP mix; 1 µl DNase I 1:250, Roche; 1 µl DNA Pol I, Thermo Scientific); the reaction is scalable up to 100 µl. The reaction was incubated at 15 °C in a waterbath for 2 h. Labeled DNA was column purified (NucleoSpin Gel and PCR Clean-up Kit, Macherey-Nagel) and eluted in 25 µl H_2O (half the reaction volume). The volume was adjusted by adding 25 µl (half the reaction volume) blue stop mix (0.1 % bromophenol blue, 0.5 % blue dextran, 0.1 M NaCl, 20 mM Tris-HCl pH 7.5).

***Tg* exon probes.** Probes to specifically label exons 2-12 and 33-47 of *Tg* were generated by standard polymerase chain reaction (PCR) using a cDNA clone (transOMIC technologies, BC111467) as template. Primer sequences are indicated in the Key Resources Table as Tg Ex 2–12 fwd, Tg Ex 2–12 rev, Tg Ex 33–47 fwd and Tg Ex 33–47 rev. 2 µg of PCR amplified DNA were labeled with fluorophores by nick translation in 100 µl reactions. Incubation time at 15 °C was reduced to 1.5 h due to the small size of the DNA fragments in order to avoid excessive digestion by DNase I.

Oligoprobes for gene subregions. Fragments for FISH probes to cover either the start of *Tg*, the first half of *Tg* intron 41 or the 3' end of *Ttn* were amplified from genomic J1 ESC wt DNA. Therefore, genomic DNA was isolated from one 6 well of J1 ESC wt cells using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions.

Primers to amplify fragments for FISH probes were designed with the HD FISH probe generator platform accessible via http://www.hdfish.eu/ (Bienko et al., 2013). Probe generation was conducted as described previously (Bienko et al., 2013) in 96 well plates (LightCycler 480 Multiwell, Roche) using a LightCycler 480 Instrument II (Roche) and KAPA

SYBR FAST (KAPA Biosystems). Primers were ordered from Eurofins Genomics (Ebersberg, Germany) in a 96 well plate format in mixed plates, with each well containing both the forward and respective reverse primer at a concentration of 5 µM. Amplification reactions (per well: 25 µl 2x KAPA SYBR FAST, 4 µl of the 5 µM forward and reverse primers dilution, 1 µl 50 ng/µl gDNA in TE, 20 µl nuclease-free H_2O) were run using the default SYBR Green program. Primer sequences are indicated in Supplementary Table S5.

After amplification in 96 well plates, fragments corresponding to one probe were mixed in a 15 ml tube and ethanol precipitated. Then, 1 µg of DNA was labeled with fluorophores using the corresponding Ulysis Nucleic Acid Labeling Kit (Invitrogen) according to the manufacturer's instructions with minor modifications. Since we experienced that longer incubation times yield a higher labeling efficiency, the labeling reaction was carried out at 80°C for 60 min instead of 15 min. We found that labeling with Alexa 647 and Alexa 594 works particularly well. We do not recommend labeling with Alexa 561, since the fluorophore was subject to severe photobleaching when imaged with confocal microscopy. Unbound dye was removed by gelfiltration using Micro Bio-Spin Columns with Bio-Gel P-30 (BioRad) or KREAPure columns (Kreatech) according to manufacturer's instructions. KREAPure columns generally yielded higher DNA concentrations after elution. All product from one labeling reaction was used to generate a FISH probe of 5 µl. The labeled product was ethanol precipitated using 10 µg of sheared salmon sperm DNA (Thermo Fisher), dried in a SpeedVac (Vacuum concentrator, Bachofer) and resuspended in 5 µl of hybridization buffer (50 % formamide, 10 % dextran sulfate, 1 x SSC).

Chromosome paints. Whole chromosome paints were a kind gift from Nigel Carter and Johannes Wienberg (Cambridge) and amplified as well as fluorophore or hapten-labeled with Biotin or Digoxigenin via DOP-PCR with a partially degenerate oligonucleotide primer (6MW primer: 5'- CCG ACT CGA GNN NNN NAT GTG G -3', Eurogentec). For each reaction, the following reagents were mixed: 5 µl 10 x PCR buffer (PerkinElmer), 4 µl 25 mM $MgCl_2$ (PerkinElmer), 5 µl 20 µM 6MW primer, 2.5 µl 2 mM dATP/dGTP/dCTP-Mix, 4 µl 1 mM dTTP, 1 µl 1 mM label-dUTP, 26.5 µl H_2O. To this mixture, 1 µl DNA and 1 µl Taq polymerase were added (total volume 50 µl). Amplification was run on a thermocycler with the following settings: 94 °C for 3 min; 20 cycles of 94 °C for 1 min, 56 °C for 1 min, 72 °C for 5 min; 72 °C for 5 min; hold at 4 °C.

Preparation of hybridization mixtures. In general, hybridization mixtures contain fluorophore- or hapten-labeled DNA as well as repeat-enriched Cot-1 DNA and salmon sperm DNA to block non-specific hybridization. All components were mixed well in a 1.5 ml reaction tube. For probe precipitation, at least 2.5 times the volume of ice-cold ethanol was added.

A typical hybridization mixture for 2D FISH with a BAC derived probe and a chromosome paint was set-up as follows: 15 µl chromosome paint–BIO, 25 µl BAC–DIG, 20 µl Cot-1 DNA (Thermo Fisher), 2 µl salmon sperm DNA (Thermo Fisher), 300 µl pre-cooled (-20 °C) EtOH. A typical hybridization mixture for 3D FISH with a BAC derived probe was set-up as follows: 100 µl BAC–fluorophore/hapten, 25 µl Cot-1 DNA (Thermo Fisher), 3 µl salmon sperm DNA (Thermo Fisher), 500 µl pre-cooled (-20 °C) EtOH. If more BACs were to be combined in the same probe, the volume of Cot-1, salmon sperm and EtOH were upscaled accordingly. A typical hybridization mixture for FISH with PCR derived probes was set-up as follows: 10 µl PCR probe–fluorophore, 10 µl salmon sperm DNA (Thermo Fisher), 250 µl pre-cooled (-20 °C) EtOH.

Mixtures were incubated for > 2 h (or overnight) at –20 °C and spun down at 14,000 rpm for 30 min at 4 °C in a tabletop centrifuge. EtOH was removed and pellets were dried in a SpeedVac (Vacuum concentrator, Bachofer) with heating for approx. 5 min. The pellets were dissolved in 2.5 µl (PCR derived probes) or 6 µl (BAC derived probe) of 100% ultrapure deionized formamide (FA, VWR Amresco). Once the pellet had dissolved, an equal volume of a master mix (20% dextran sulfate in 2xSSC) was added to the probe and mixed by vortexing. Hybridization mixtures were kept at –20°C.

2D-FISH. 2D-FISH experiments were carried out on metaphase spreads of mouse embryonic fibroblasts (MEFs) to verify the correct position of the selected BACs on the respective chromosome. BACs were labeled with digoxigenin-dUTP via NT and chromosome paints were labeled with biotin-dUTP via DOP-PCR. Probes were loaded on metaphases under 15 x 15 mm coverslip and sealed with rubber cement (Fixogum) to prevent drying during hybridization. DNA of probes and chromosomes was denatured simultaneously on a heat block at 75 °C for 3 min.

Hybridization was carried out in a dark chamber in the waterbath at 37 °C for 1 or two days. Post hybridization washings were carried out using 2 x SSC at 37 °C (3 x 10 min) and 0.1 x SSC at 60 °C (1 x 5 min). Detection was carried out using streptavidin-Alexa 555 (Molecular Probes, dilution 1:200) and sheep-anti-digoxigenin antibody (Roche, dilution 1:200), dissolved in blocking solution (4 % BSA/4 x SSCT), for 1 h. Washings were performed using 4 x SSCT (2 x 10 min). DNA was counterstained with 2 µg/ml DAPI in 2 x SSC (Sigma-Aldrich) for 10 min. Slides were rinsed with 2 x SSC and then thoroughly washed with distilled H_2O, air dried in darkness and mounted in Vectashield (Vector Laboratories). The edges of the coverslip were sealed with nail polish.

Subculturing and fixation of cells for FISH or immunostaining. 15 x 15 mm square coverslips (R. Langenbrinck or Carl Roth) were thoroughly cleaned with 80 % ethanol, coated with poly-L-lysine (Sigma-Aldrich) for 15 min and sterilized in 12-well plates in a tissue culture hood under UV light for 1 h before cell seeding. Coverslips with cells were washed twice with pre-warmed (37°C) cell culture grade PBS (Sigma-Aldrich) and fixed with 4 % PFA (Merck)/PBS for 10 min at RT. During the last minute of fixation, a drop of PBS/0.5 % Triton X-100 was added to prevent cell distortion during changes of solutions. Cells were washed 3x with PBS/0.01 % Triton X-100 for 5 min and then permeabilized with PBS/0.5 % Triton X-100 for 10 min.

FISH and immuno-FISH on cultured cells. FISH on cells was performed as previously described (Solovei and Cremer, 2010). Coverslips with fixed cells were incubated with 20% glycerol/PBS for 1 h and subjected to 4-5 rounds of freezing/thawing using liquid nitrogen. Then, cells were washed three times with PBS/0.01 % Triton X-100 for 10 min, rinsed in 2 x SSC and equilibrated in 50 % FA/2 x SSC. For DNA FISH, before equilibrating in 50 % FA/2 x SSC, cells were treated with RNase to ensure hybridization only to DNA. For this coverslips were briefly washed in PBS, placed up-side-down on drops of 50 µg/ml RNase on a piece of Parafilm, incubated at 37°C for 1 h and washed in PBS (3 x 10 min). For hybridization, a coverslip with cells was carefully placed up-side-down on a top of a probe drop loaded on a microscopic slide and the edges of the coverslip were sealed with rubber cement (Fixogum). For DNA-FISH or FISH detecting both DNA and RNA, denaturation of both probe and sample DNA was carried out simultaneously on a hot block at 75°C for 3 min. For RNA-FISH, only the probe DNA, but not the sample DNA was denatured.

Hybridization was carried out in metal boxes floating in a water bath at 37°C for 2 days. Post hybridization washings included 2 x SSC at 37°C (3 x 30 min) and 0.1 x SSC at 60°C (1 x 7 min). In case of probe labeling with haptens (biotin, digoxigenin) detection was carried out as described for 2D FISH. DNA was counterstained with 2 µg/ml DAPI (Sigma-Aldrich) for 10 min, cells were mounted in Vectashield (Vector Laboratories) and the edges of the coverslip were sealed with nail polish.

Immuno-FISH was performed as previously described (Solovei and Cremer, 2010). Briefly, incubations with primary and secondary antibodies diluted in blocking solution (PBS/4 % BSA/0.01 % Tween 20) were carried out for 1 h each. Washings in between incubations and after the secondary antibody were performed with PBS/0.1% Tween 20 (3 x 15 min). Cells were post-fixed with 2% PFA/PBS for 10 min, incubated in 20% glycerol for 1h and further treated accordingly to the standard protocol for 3D-FISH.

Fixation and sectioning of tissues. CD1 mice (Charles River Laboratories) were sacrificed by cervical dislocation after anesthetizing with IsoFlo (Isofluran, Abbott). For isolation of the thyroid gland, a vertical anterior neck incision was made. After removing the parotid glands and surrounding muscle tissue, the thyroid lobes were visible below the thyroid cartilage on each side of the trachea. A piece of trachea (ca. 5 mm) with the attached thyroid lobes was excised and placed into PBS. Under a binocular, the thyroid tissue was carefully cleaned from connective and fat tissue, muscles and parathyroid glands using fine forceps and fine scissors (Fine Science Tools).

In order to relax smooth muscles of a bladder, mice were injected with Narcoren (Boehringer Ingelheim, conc. 5 µl/g body weight). The abdominal musculature was carefully removed, the bladder was excised and placed into PBS in a small plastic dish. For isolation of colon, a vertical incision in the lower abdomen was made and extended horizontally in both directions. The abdominal musculature was carefully removed and a part of the distal colon was excised.

All tissues were washed once with PBS and then fixed in 4 % paraformaldehyde (Carl Roth) solution in PBS for 12-20 h. After fixation, mouse tissues were washed in PBS and incubated in increasing concentrations of sucrose: 10 % for 1 h, 20 % for 1 h and 30 % for 1 to 12 h. For embedding, tissues were placed into flexible silicone rubber mold (Pelco) filled up with Tissue-Tek O.C.T. compound freezing medium (Sakura) in the desirable orientation. Once the medium with tissue was frozen, blocks were stored in boxes at –80°C. Frozen mouse tissues were cut into 16-20 µm sections using a cryostat (Leica CM3050S) on Superfrost Plus slides (Thermo Scientific) and stored at –80°C before use.

For preparation of thin sections (50-70 nm), thyroid glands were fixed with mixture of 2 % paraformaldehyde and 0.1 % glutaraldehyde in 300 mOsm cacodylate buffer (75 mM cacodylate, 75 mM NaCl, 2 mM $MgCl_2$) for 30 min and immediately embedded into Lowicryl.

FISH and immuno-FISH on cryosections. FISH on cryosections was performed as previously described (Solovei, 2010). Briefly, cryosections were removed from -80°C freezer and dried for 20 to 30 min at RT. Slides were then placed into a 10 mM sodium citrate buffer (pH 6.0) for 3 min for rehydration, heated up to +80°C for 20 min in the same buffer for antigen retrieval, and then equilibrated in 2 x SSC. For DNA-FISH, after this step sections were treated with RNase as described above for cells and subsequent washings in PBS were carried out for 3 x 30 min. Then slides were incubated in 50 % formamide/ 2 x SSC for 30-60 min. For DNA-FISH a probe was loaded under the chamber prepared from coverslips (Solovei, 2010), the edges of the chamber were sealed with rubber cement and both probe and sample DNA were denatured simultaneously on a hot block at +80°C for 3 min. For RNA-FISH, only the probe DNA, but not the sample DNA was denatured. Hybridization was carried out in metal boxes floating in a water bath at 37°C for 2 days and post hybridization washings included 2 x SSC at 37 °C (3 x 30 min) and 0.1 x SSC at 60 °C (1 x 7 min). In case of probe labeling with haptens (biotin, digoxigenin) detection was carried out as described for 2D FISH. DNA was counterstained with 2 µg/ml DAPI (Sigma-Aldrich) in PBS for 30-60 min, washed with PBS for 30 min. Sections were mounted in Vectashield (Vector Laboratories) and the edges of the coverslip were sealed with nail polish.

For immuno-FISH, the FISH protocol on cryosections was followed and after the post-hybridization washings, slides were equilibrated in PBS, permeabilized with 0.1 % Triton X-100/PBS for 30 min and incubated with primary and then secondary antibodies diluted in blocking solution (1% BSA/0.1% Triton X 100/0.1% saponin). Incubations with both antibodies were carried out overnight at RT under a glass chamber in a dark humidified chamber. DAPI (2 µg/ml) was added directly to the secondary antibody. Washings between antibodies and after the secondary antibody were performed with 0.01 % Triton X-100/PBS at 37 °C (3 x 30 min).

Microscopy and image analysis

Confocal laser scanning microscopy. Image stacks were acquired using a TCS SP5 confocal microscope (Leica) using a Plan Apo 63/1.4 NA oil immersion objective and the Leica Application Suite Advanced Fluorescence (LAS AF) Software (Leica). Z step size was adjusted to an axial chromatic shift and typically was either 200 nm or 300 nm. XY pixel size varied from 20 to 60 nm, depending on the final magnification. Used laser lines: 405 nm diode (DAPI), 488 nm Argon (FITC, Alexa 488), 561 nm DPSS (Cy3, TAMRA, Alexa 555), 594 nm HeNe (Texas Red, Alexa 594) and 633 nm HeNe (Cy5, Alexa 647). Obtained confocal stacks were processed using the ImageJ software (https://imagej.nih.gov/ij/). Before analysis, stacks were corrected for chromatic shift in z direction using the StackGroom/z-shift corrector plugin and RGB image stacks, montages of single optical sections and maximum intensity projections were generated using the StackGroom/3channels plugin (Walter et al., 2006). Images in the manuscript are mostly maximum intensity projections of several optical sections of a z-stack, unless stated otherwise.

Semi-automated measurements of inter-flank distances. The signal spots in both channels were identified by detecting local minima in a Laplacian-of-Gaussian (LoG)-filtered image (with the expected size/sigmas set to enhance spots at the diffraction limit). A pairing of spots from both channels with minimal total distance was calculated using linear assignment and coordinates of both partners saved for further analysis. In both the detection and matching steps, results were visualized immediately, with the option to manually curate them and remove erroneous detection of pairings. The spot pair detection was implemented in Python in the form of a Jupyter notebook. The code is available at https://github.com/hoerldavid/fish_analysis.

Measurement of TL compaction rate. For measuring the contour length of the *Tg* TL regions, maximum intensity projections of stacks after RNA-FISH were used. Loop contours were manually traced using the *Segmented Line tool* in ImageJ.

Measurements of TL signal size. For comparison of RNA-FISH signal size of *Cald1* TLs in control cells and after splicing inhibition, maximum intensity projections of stacks after RNA-FISH were prepared and filtered (Gaussian = 1). The signal areas of individual *Cald1* alleles were measured in Fiji after applying *Intermodes* threshold.

Transcription loop modeling

Modeling parameters. We modeled chromatin as a polymer with each monomer representing 1 kb or roughly 5 nucleosomes. To fix the three-dimensional (3D) size of a monomer we used literature values for the compaction ratio of the chromatin fiber, i.e. the number of bp per nm of contour length. Literature values vary greatly and are in the range c = 25 to 150 bp/nm (Bystricky et al., 2004; Dekker et al., 2002; Rippe, 2001). We here used c = 50 bp/nm, which fixed the 3D size or our 1kb monomers to 20 nm.

For the Kuhn length lK (typical bending radius of the fiber), again, literature values vary greatly and are in the range of lK= 60 … 400 nm (refs as above). Further estimations of the Kuhn length come from our own measurements of spatial distances of the 5' and 3' flanks of our genes in the inactive state (see …). Those measurements suggest somewhat smaller Kuhn lengths between 25 and 80 nm. For our polymer model, we here chose lK= 56 nm or 2.8 kb.

Finally, we fixed the volume density of chromatin in the cell nucleus, i.e. what fraction of the nucleus is occupied by the polymer fiber. A proxy can be obtained from the genome size (≈ 3 Gb for a single copy, ≈ 6 Gb per nucleus) and the size of the nucleus. With a diameter of 8 um we get a volume density of 180 kb in a cube of size (20 nm)^3. A cryo-EM measurement (Ou et al., 2017) suggests volume densities around 30%. For our polymer simulations we chose = 0.18 beads / (20 nm)^3 or = 20%.

In order to interpret simulations we converted the simulation time scale to physical time. This conversion is done by comparing mean square displacements of simulation monomers with experimental data on tracked loci in yeast (Hajjoul et al., 2013) and mammals (Lucas et al., 2014) as described in (Nuebler et al., 2018). We obtain a conversion factor of 656 ps in simulations corresponding to 1s real time.

Polymer simulation software. Polymer simulations were performed using a Mirny lab written wrapper (available at https://github.com/mirnylab/openmm-polymer-legacy) around the open source GPU-assisted molecular dynamics package Openmm (1,2). Polymers are represented as a chain of monomers with harmonic bonds, a repulsive excluded volume potential, and an additional small attraction for the interaction of two monomers of type B. Furthermore, the association of B-type chromatin with the nuclear periphery was modeled by a weak attraction of B-type monomers with the sphere boundary representing the nuclear envelope.

Simulation setup. We simulated six chromosomes of 50 Mb each (corresponding 50,000 monomers each) and placed them inside a sphere of diameter of 2.84 um (corresponding to 42 monomer diameters). We generated territorial chromosomes by initiating them in a mitotic-like conformation and letting them to expand. In a dense environment, polymer dynamics is exceedingly slow, therefore, chromosomes mix only moderately and retain their territoriality over all simulated times, corresponding to at least hours, or days of real time. We repeated the above expansion from mitosis 10 times for 25 min and 24 s (1,000,000 simulation time units) each. These chromosome conformations served as initial conditions for our simulations of long highly transcribed genes.

To simulate the effect of a high expression of a long gene, we assigned a 100 kb region on each simulated chromosome to our gene of interest. In these regions the bending stiffness of the fiber was increased. The “genes” were positioned in a region of A-type chromatin. Unless otherwise indicated, simulations of highly transcribed genes were run for 127 min of real time (5,000,000 simulation time units). In each run, 101 conformations were saved at equidistant time points.

Simulation data analysis. To obtain Hi-C maps from simulated data we first coarse grained polymer conformations by a factor of 10 (i.e. only every 10th monomer is considered) in order to reduce the size of the computed Hi-C matrix. We then defined a cutoff radius, mimicking the crosslinking radius in an actual Hi-C experiment. The cutoff radius was 10 monomer diameters (we verified that results are insensitive to the cutoff). Hi-C maps were computed from conformations in the second half of each run (we verified that this choice had no effect on results). *Cis/trans* contact frequency ratio tracks were computed as for experimental Hi-C matrices. The *cis/trans* ratio tracks of all six simulated chromosomes were aggregated to improve statistics.

4. QUANTIFICATION AND STATISTICAL ANALYSIS

Explanations of statistical analyses, error bars, parameters and quantifications in this study are described in the main and supplemental text and figure legends. All of the statistical tests were performed by using the wilcox.test function implemented in R. P-values are indicated in the respective figures and figure legends.