TRANSCRIPTION SHAPES GENOME FOLDING ON A SINGLE GENE LEVEL

Dissertation an der Fakultät für Biologie der

Ludwig-Maximilians-Universität München

vorgelegt von

Simon Ullrich

München, 2024

Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Dr. Heinrich Leonhardt
am Lehrstuhl für Humanbiologie und BioImaging
an der Ludwig-Maximilians-Universität München

Erstgutachter:	Prof. Dr. Heinrich Leonhardt
Zweitgutachter:	Prof. Dr. Peter Becker
Tag der Abgabe:	24.07.2024
Tag der mündlichen Prüfung:	05.12.2024

Table of Contents

List of Figures

Abstract

This thesis aims to investigate the spatial arrangement and structural characteristics of transcribed genes during their active expression in eukaryotic cells. Given the limited resolution of light microscopy, we focused on long and highly expressed genes. Our findings revealed that these genes undergo significant reorganization within their loci, expanding from chromosomes in the form of open loops, which are densely decorated with elongating RNA polymerases (RNAPIIs) and the attached nascent ribonucleoproteins (RNPs). We have designated these structures as transcription loops (TLs) and demonstrated that RNAPIIs move along genes during transcriptional elongation, carrying nascent RNAs with them. This challenges the conventional model of transcription factories and instead suggests a universal mechanism of eukaryotic transcription (Leidescher et al., 2022), which was assumed in the early studies on giant polytene and lampbrush chromosomes.

Furthermore, to reconcile our observations with the foci-like patterns of transcription sites observed with small genes, we proposed the following hypothesis: The extension of long, highly expressed genes is driven by their intrinsic stiffness, resulting from the dense decoration of the gene with RNAPIIs and voluminous nascent RNPs. This hypothesis was successfully tested, revealing that such a dense decoration occurs only at very high levels of transcription and that voluminous RNPs can be formed over long introns or exons. This explains why we only detect TLs formed by long, highly upregulated genes. Small genes lacking long introns or exons are decorated with much smaller RNPs, allowing them to loop within a smaller volume, leading to structures unresolvable by light microscopy (Leidescher et al., 2022).

One of the most extensively studied TLs in our research is formed by the Thyroglobulin (*Tg*) gene. This gene is characterized by high upregulation in the thyroid gland across various vertebrate groups, from fish to mammals, as well as in mouse and human organoids. The TG protein serves as a scaffold for the storage and production of T3/T4 thyroid hormones. We demonstrated that *Tg* transcription remains consistently elevated, regardless of thyroid hormone levels, circadian rhythmicity, or the exocrine versus endocrine phases of thyrocyte activity. By establishing the *Tg* gene as a unique model of sustained high transcriptional activity, our study provides a valuable system to investigate robustly upregulated eukaryotic gene expression *in vivo*. This includes intriguing questions such as exploring the role of enhancer-promoter contacts for maintaining the high upregulation of expression of genes (Ullrich et al., 2024).

TLs formed by the studied lengthy genes facilitated our research of several aspects of transcription using fluorescence microscopy techniques at the single-cell level. We observed a disproportionate size of intron signals compared to entire *Tg* TLs and demonstrated that complete splicing of *Tg* introns is delayed by approximately 50-60 kb downstream of the transcription site of an

intron. This may be due to local exhaustion of the splicing machinery caused by high *Tg* transcriptional activity. It is paradoxical that the longest *Tg* intron, 50 kb in length, exhibited much quicker dynamics of splicing, similar to the promptly excised long introns of the *Cald1* gene (Ullrich et al., BioRxiv). Ultimately, our analyses promise to unveil novel insights into the spatial organization and regulatory mechanisms governing the transcription of highly expressed genes.

Zusammenfassung

Ein zentrales Ziel dieser Arbeit ist es, die räumliche Anordnung und die strukturellen Eigenschaften von transkribierten Genen während ihrer aktiven Expression in eukaryotischen Zellen zu untersuchen. Unter Berücksichtigung der begrenzten Auflösung der Lichtmikroskopie konzentrierten wir uns auf lange und hoch exprimierte Gene und können feststellten, dass sie ihren chromosomalen Locus stark reorganisieren und sich von den Chromosomen in Form von Transkriptionsschleifen abheben. Diese sind dicht mit verlängernden RNA Polymerasen (RNAPIIs) und angehängten naszierenden Ribonukleoproteinen (RNPs) bedeckt. Wir haben diese Strukturen Transkriptionsschleifen genannt und konnten zeigen, dass sich RNAPIIs während der Transkription entlang von Genen bewegen und naszierende RNAs mit sich führen. Damit stellen unsere Ergebnisse das Modell der Transkriptionsfabriken in Frage und legen nahe, dass es sich um einen universellen Mechanismus der eukaryotischen Transkription handelt, der auch in frühen Studien an Riesenpolytän- und Lampenbürstenchromosomen beschrieben wurde.

Um unsere Ergebnisse mit den punktartigen Mustern der Transkriptionsstellen, die von kleinen Genen erzeugt werden, in Einklang zu bringen, haben wir die folgende Hypothese aufgestellt: Die Ausdehnung eines langen, hochexprimierten Gens wird durch die Steifheit verursacht, die sich aus der dichten Besiedlung der Genachse mit RNAPIIs und voluminösen naszierenden RNPs ergibt. Wir haben diese Hypothese erfolgreich getestet und gezeigt, dass eine solche dichte Dekoration des Gens bei hohen Transkriptionsleveln möglich ist und dass voluminöse RNPs vorwiegend über langen Introns oder Exons gebildet werden können. Dies erklärt, warum wir TLs von langen, hochregulierten Genen nachweisen. Kleine Gene ohne lange Introns oder Exons sind mit viel kleineren RNPs dekoriert, so dass sie sich nur in einem kleineren Volumen entfalten können. Dies führt zu kleinen Strukturen, die sich durch Lichtmikroskopie nicht auflösen lassen (Leidescher et al., 2022).

Eine der in unserer Forschung am intensivsten untersuchten TLs wird durch das Thyreoglobulin (*Tg*) Gen gebildet. Dieses Gen zeichnet sich durch eine starke Hochregulierung in der Schilddrüse in verschiedenen Wirbeltieren, von Fischen bis hin zu Säugetieren, sowie in Organoiden von Mäusen und Menschen aus. Das TG-Protein dient als Grundgerüst für die Speicherung und Produktion von T3/T4-Schilddrüsenhormonen. Wir konnten nachweisen, dass die *Tg*-Transkription unabhängig vom Schilddrüsenhormonspiegel, der zirkadianen Rhythmik oder den exokrinen bzw.

endokrinen Phasen der Thyreozytenaktivität konstant erhöht bleibt. Durch die Etablierung des *Tg*-Gens als einzigartiges Modell für eine anhaltend hohe Transkriptionsaktivität bietet unsere Arbeit ein wertvolles System zur Untersuchung der robust hochregulierten eukaryotischen Genexpression *in vivo*. Dazu gehört beispielsweise auch die Notwendigkeit von Enhancer-Promoter-Kontakten zur Aufrechterhaltung einer hohen Hochregulierung von Genen (Ullrich et al., 2024).

Die TLs der untersuchten langen Gene helfen uns verschiedene Aspekte der Transkription mit Hilfe von Fluoreszenzmikroskopie auf Einzelzellebene zu untersuchen. Wir haben eine unverhältnismäßige Größe der Intron-Signale im Vergleich zu den gesamten *Tg*-TLs nachgewiesen und konnten zeigen, dass das vollständige Spleißen von *Tg*-Introns um ca. 50-60 kb nach Transkription der jeweiligen Introns verzögert stattfindet. Dies führen wir auf die hohe *Tg*-Transkriptionsaktivität zurück, die die Spleißmaschinerie mutmaßlich lokal erschöpft. Paradoxerweise zeigt das längste *Tg*-Intron mit einer Länge von 50 kb eine viel schnellere Spleißdynamik, ähnlich wie bei den schnell gespleißten langen Introns des *Cald1*-Gens (Ullrich et al, BioRxiv). Unsere Analysen bieten neue Einblicke in die räumliche Organisation und die regulatorischen Mechanismen, die die Transkription hochexprimierter Gene steuern.

1 Introduction

Chromatin serves as the dynamic scaffold for organizing and packaging the genome, thereby regulating various cellular processes such as gene expression, DNA replication, and repair. Chromatin provides the spatial and functional framework, which is necessary for the precise regulation of cellular activities and ultimately determines a cell's fate and function. The study of chromatin and its structure has undergone significant advancements over the years, most of which were driven by technological innovations and novel interdisciplinary approaches. Starting with rudimentary observations enabled by basic light microscopy techniques, the field has progressed to utilize innovative methods such as high-resolution imaging, chromosome conformation capture and modern sequencing techniques, alongside single-cell analysis, as well as molecular tools for genome manipulations. This evolution has transformed the study of chromatin structure into a multifaceted discipline at the forefront of modern biology.

1.1 Chromatin

1.1.1 Chromatin structure

At its most basic level, chromatin consists of DNA wrapped around histone proteins, which together form the so-called nucleosomes. Each nucleosome core particle consists of around 147 base pairs of DNA wrapped 1.7 times around an octamer histone complex including two copies of each histone protein H2A, H2B, H3, and H4 (Luger et al., 1997). Individual nucleosomes are arranged in a linear fashion with regularly spaced linker DNA sequences between, ranging from 10 to 50 base pairs in length. Nucleosome appearance was initially described as "beads on a string" with an approximate diameter of 11 nanometers (Baldi et al., 2020; Olins et al., 1992; Olins and Olins, 2003). The linker histone H1 connects adjacent nucleosomes and helps with stabilizing and compacting the chromatin fiber by binding to the linker-DNA segments (Allan et al., 1980). Post-translational modifications of histone proteins and chemical modifications of DNA act as regulatory mechanisms that modulate chromatin structure and accessibility and thus represent the epigenetic level of chromatin regulation (Waddington, 1942).

1.1.2 Functional chromatin compartmentation into eu- and heterochromatin

Functionally, the chromatin within an interphase nucleus can be broadly categorized into two types based on structural and functional criteria: euchromatin (EC) and heterochromatin (HC) (Heitz, 1928). EC is generally less condensed, located to the interior of the nuclear space, contains most transcribed genes and as such is accessible for proteins that regulate transcription, such as transcription factors (TFs). EC is characterized by higher CpG content as well as hyper-acetylation of core histones H3 and

H4 along with tri-methylation of H3K4 (van Steensel and Belmont, 2017). In contrast, HC forms a more compact structure, is typically devoid of genes, less accessible to proteins and is often concentrated near the nuclear envelope and the nucleolus (Solovei et al., 2016). As such, HC exhibits only low transcriptional activity.

HC is roughly divided into two parts, facultative and constitutive. Constitutive HC predominantly consists of highly repetitive DNA sequences such as satellite DNA and remains condensed throughout most of the cell cycle and regardless of the cell type. A typical mark of constitutive HC is the tri-methylation of H3K9 and H4K20, which act as a binding platform for the recruitment of HP1 (Heterochromatin Protein 1) which plays multiple roles in propagating and maintaining the heterochromatic state (Bannister et al., 2001; Peters et al., 2003; Schotta et al., 2004). Facultative HC contains tissue-specific and differentially expressed genes that can be rapidly up- or downregulated at certain developmental or differentiation stages. An example for facultative heterochromatin is the inactivation of one X chromosome in mammalian female cells (Dossin et al., 2020; Trojer and Reinberg, 2007). Facultative heterochromatin is marked by the repressive histone modifications like H3K27me3, H2AK119Ub, which are mediated by polycomb repressor complexes 1 and 2, respectively, as well as by H3K9me2 and H4K20me3 (Heard, 2005; Wang et al., 2004).

1.1.3 Chromatin mobility

Whereas on a large scale, nuclear chromatin is largely immobile, on a fine scale, it is not static but exhibits dynamic properties. The composition and positioning of nucleosomes, along with the distribution of epigenetic marks, can undergo frequent alterations. Moreover, many interactions between chromatin and nuclear proteins, such as TFs or heterochromatin-binding proteins, are highly transient in nature (Voss and Hager, 2014). Especially the interactions between transcription factors and chromatin are characterized by mutual regulation and interdependence. On one hand, TFs must initiate local chromatin remodeling and decompaction to bind to their target sites and facilitate transcriptional activity. Conversely, chromatin compaction plays a regulatory role in TF binding, as regions of heterochromatin render target sites inaccessible to TFs. Additionally, chromatin compaction does regulate the access of replication-initiation factors to alternate compartments, thereby also regulating the timing of DNA replication (Bulger and Groudine, 2011; De Laat and Duboule, 2013).

Another aspect of chromatin dynamics is connected to the local nanoscale movements of genomic regions, which are reflected by their transcriptional activity. The positioning of genes is regulated by changes in their transcriptional status, as silenced genes are typically associated with heterochromatin (HC), while active genes are associated with euchromatin (EC). The major universal compartment for gene silencing is the peripheral heterochromatin (Poleshko and Katz, 2014). Every

chromosome includes genomic segments that bind to the nuclear envelope, forming structures known as lamina-associated domains (LADs) (Guelen et al., 2008; Pickersgill et al., 2006). Gene activity is locally regulated through the binding and release of LADs, although the causal relationship between these processes remains unclear (Peric-Hupkes et al., 2010). Some evidence suggests that genes are released from the peripheral heterochromatin before they become active (See et al., 2019), while other data indicate the opposite sequence of events (Leemans et al., 2019).

1.2 Chromatin Domains

1.2.1 Spatial segregation of eu- and heterochromatin and compartmentalization into A/B compartments

Originating from the chromatin fiber, chromatin undergoes intricate organizational hierarchies to achieve a higher-order 3D structure shaping the functional landscape of the nucleus (Bickmore, 2013). The first hierarchical level is the segregation of EC and HC in the nucleus, as discussed above. Microscopic investigations have further unveiled the spatial segregation of HC and EC within the nucleus: HC predominantly occupies regions at the nuclear periphery and around the nucleoli, while EC is situated in the interior of the nucleus (Goetze et al., 2007; Kupper et al., 2007). The utilization of chromosome-conformation-capture techniques yields a genome contact matrix that quantifies the level of interaction between two loci in the genome. Based on this, chromatin can be categorized into distinct nuclear neighborhoods termed compartments. Active chromatin regions predominantly occupy the A compartment, while inactive chromatin regions predominantly occupy the B compartment. These compartments closely correspond to EC and HC, respectively. Loci situated within the same compartment exhibit relatively frequent contacts, even when they are positioned at considerable distances along a chromosome or on entirely different chromosomes. (Lieberman-Aiden et al., 2009; Rao et al., 2014; Zhang et al., 2012). Extensive alterations in genome compartmentation patterns can occur between different cell types, cellular conditions or during differentiation, and these changes are pivotal in both the activation and repression of genes, contributing significantly to the establishment and maintenance of cell-type-specific gene expression programs (Brero et al., 2005; Dixon et al., 2015).

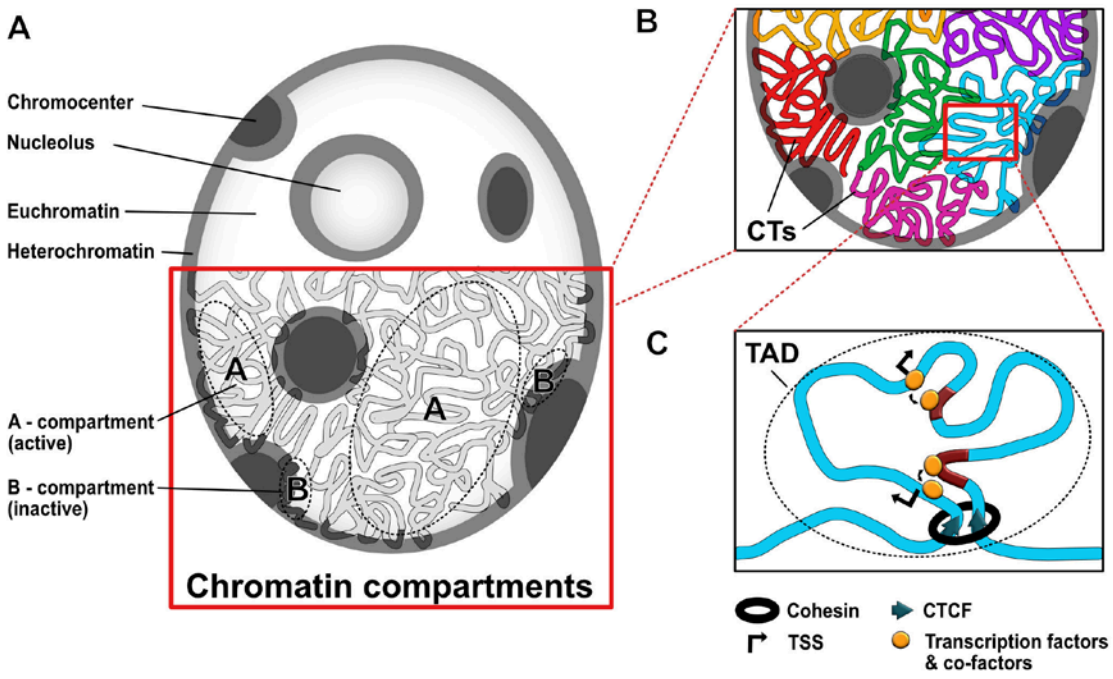


Figure 1: Hierarchical levels of genome organization within the nucleus.

(A) Active euchromatin (light gray) in the nuclear center is spatially separated from inactive heterochromatin (medium gray) at the nuclear periphery as well as around nucleoli (light gray) and chromocenters (dark gray) (upper part). Chromatin is divided into two main compartments: active and open A and inactive and closed B compartments (lower part), representing the euchromatic and heterochromatic regions, respectively. **(B)** Individual chromosomes (highlighted in yellow, red, green, cyan, magenta and pink) occupy discrete chromosome territories (CTs) and traverse between the two compartments, depending on the transcriptional activity of the respective region. **(C)** Topologically associating domains (TADs) are defined by regions with elevated long-range interaction frequencies, such as promoter-enhancer loops, and are delimited by boundaries marked by architectural proteins like cohesin and CTCF.

1.2.2 Chromosome territories (CT)

Chromosomal territoriality stands as a cornerstone paradigm in nuclear biology (Cremer and Cremer, 2010). This paradigm, in its modern iteration, emerged in the 1980s, contradicting a perspective originating from electron microscopy, which suggested that chromosomes were dispersed and intertwined throughout the nucleus (Cremer and Cremer, 2006). However, several observations have raised questions regarding the functional significance of territoriality, hinting instead that chromosome territories (CTs) may merely result from the latest mitosis in conjunction with the relative immobility of interphase chromatin (Branco and Pombo, 2006; Rosa and Everaers, 2008). Following mitosis, the chromosomes decondense into seamless chromatin lacking distinguishable individual chromosomes. Nonetheless, visualization using paint probes reveals the individual chromosomes localization within the nucleus, creating the impression of "territoriality" (Cremer and Cremer, 2010). However, balanced chromosome rearrangements, which do not disrupt genes or regulatory elements, do not influence nuclear functions, despite altering the landscape of chromosome territories

(Warburton, 1991). Similarly, strong karyotype reorganization between closely related species, such as the difference between the Chinese muntjac and the Indian muntjac (Macgregor, 1993) further illustrates the existence of markedly distinct chromosome territories without affecting cellular or organismal physiology.

Subsequent to the introduction of the term "Chromosome Territory," considerable attention has been devoted to studying the shape and positioning of CTs in various cell types and conditions. Nowadays however, the uneven distribution of genic segments and LADs is sufficient to explain the positioning and shape of CTs. Indeed, the abundance and distribution of both LADs and genes along a chromosome are intricately interconnected (Guelen et al., 2008; Solovei et al., 2016). Gene-poor chromosomes with abundant LADs are tethered to the nuclear envelope, adopting a flattened shape and remaining at the nuclear periphery. In contrast, gene-rich chromosomes, bound to the nuclear periphery primarily by the pericentromeric region, tend to assume an internal position due to the absence of multiple LADs. This phenomenon is particularly evident in large nuclei, such as those found in neurons or stem cells (Solovei et al., 2016).

1.2.3 Topologically associating domains (TADs)

Advancements in chromosome-conformation-capture techniques such as 5C- and Hi-C have unveiled more structural details of interphase chromosomes. It has been discovered that on a smaller scale chromatin is arranged into topologically associating domains (TADs) spanning up to hundreds of kilobases (Dixon et al., 2012; Galupa and Heard, 2017). TADs consist of continuous areas of chromatin with a 2-3 times elevated contact frequency compared to chromatin contacts between different TADs. The formation of TADs has been attributed to chromatin loop extrusion by the cohesin complex which reels chromatin into a loop from both sides, a process, which can be blocked by the chromatin-bound CCCTC-binding factor (CTCF), and thus TADs are mostly formed between convergent CTCF sites (de Wit et al., 2015; Guo et al., 2015; Kim et al., 2019). Importantly, this blocking of chromatin extrusion by CTCFs partitions the genome into partially insulated neighborhoods. It is considered that genes residing within the same TAD are predominantly co-regulated (Dowen et al., 2014). Disrupting TAD insulation results in significant misregulation of gene expression, as gene promoters can come into contact with erroneous regulatory elements, leading to developmental abnormalities and cancer (Flavahan et al., 2016; Hnisz et al., 2016; Lupianez et al., 2015).

1.3 The Spatial Organization of Transcription

Transcription greatly influences chromatin structure through a complex interplay of molecular interactions. Upon activation, chromatin remodeling complexes are recruited to specific genomic regions by TFs, facilitating the repositioning or eviction of nucleosomes and creating open chromatin

configurations conducive to transcription initiation and elongation (Brown, 1984; Steger and Workman, 1996; Workman, 2006). Furthermore, the process of transcriptional elongation by RNA polymerase II (RNAPII) involves the transient displacement and subsequent reassembly of nucleosomes along the DNA template, contributing to the dynamic modulation of chromatin structure (Kristjuhan and Svejstrup, 2004; Lee et al., 2004). Transcription also induces long-range chromatin interactions, bringing distant regulatory elements into spatial proximity (Hampsey et al., 2011). Overall, transcriptional processes play a pivotal role in shaping chromatin architecture, contributing to the dynamic regulation of gene expression and cellular function (Friman and Bickmore, 2022).

1.3.1 Early views on the spatial transcription organization

An early key contribution to our view on individual transcribed genes has emerged from Miller spreads. Miller spreads involve treating chromatin with very low ionic strength and alkaline pH, allowing its gentle unfolding and spreading onto EM grids. After staining, chromatin spreads on grids are examined under an electron microscope (Miller and Beatty, 1969). Miller spreads have been particularly valuable for understanding the morphology of transcribed genes and nascent RNPs (Figure 2 A), but at the same time they were lacking comprehension of the spatial organization of transcribed genes in their natural context. More information came from studying gigantic chromosomes, such as lampbrush chromosomes (LBCs) isolated from diplotene oocytes of amphibians, and polytene chromosomes of Diptera.

LBCs are diplotene bivalents formed during meiosis in the oocytes of all vertebrates apart from mammals (Macgregor, 1993). On these chromosomes, transcription units, as a rule surpassing the length of individual genes, exhibit extensive transcriptional activity, which results in the formation of chromatin loops extending laterally from the chromatid axis (Figure 2 B). This phenomenon gives rise to the characteristic appearance of bivalents resembling fluffy brushes (Callan, 1986).

Polytene chromosomes, derived from consecutive chromatid duplications in the absence of cell division, comprise numerous homologous DNA strands tightly bound together. They are present in the interphase nuclei of various fly tissues, notably prominent in larval salivary glands (Bjork and Wieslander, 2015). Similar to lampbrush chromosomes, actively transcribed genes extend outward from the chromosome axis. However, due to the presence of thousands of adjacent strands, these genes do not form individual loops but rather numerous collateral loops referred to, according to their appearance, as "puffs". The largest ones are known as "Balbiani rings" (Figure 2 C) (Daneholt, 1975).

Importantly, for both LBCs and polytene chromosomes, the spatial positioning of the 5’ and 3’ ends of the loops has been shown to remain stationary, while RNAPIIs progressively move and transcribe along the DNA template, transporting increasingly elongated nascent RNA transcripts. LBC loops and polytene puffs share the considerable size of their studied transcription units as well as their

notably high expression levels, which are the key features facilitating their visualization by microscopy (Bjork and Wieslander, 2015; Macgregor, 1993).

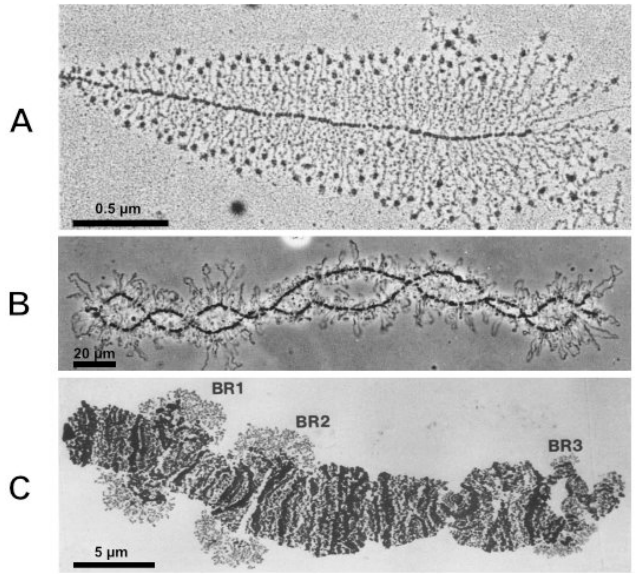


Figure 2: Models for the early microscopic visualization of transcription.

(A) Miller spreads of rRNA genes. The flattened gene axis is visible, with nascent RNAs emerging from it. As transcription progresses from left to right, RNAs increase in size, resulting in the characteristic "Christmas-tree" shape associated with this type of chromatin preparation. Adapted from (Scheer, 1987). **(B)** Phase-contrast micrograph of a lampbrush chromosome from a *Pleurodeles waltlii* oocyte. Bivalent chromosome axes are covered by lateral loops formed by highly expressed transcription units. Adapted from (Scheer, 1987). **(C)** An electron micrograph of isolated chromosome IV *of Chironomus tentan*s. Highlighted are the three prominent puffs known as Balbiani rings (BR) 1, 2, and 3. Adapted from (Björkroth et al., 1988).

1.3.2 The “transcription factory” model

Decades later, when the focus of cytogenetics shifted towards interphase nuclei, several microscopic observations of actively transcribed genes, elongating polymerases, and nascent RNA transcripts, led to the concept of the “transcription factory” (Cook, 1999; Ghamari et al., 2013; Papantonis and Cook, 2013). In contrast to earlier perspectives on transcription, which posited the stationary nature of the DNA template and the movement of RNAPIIs along a gene, carrying a nascent ribonucleoprotein (nRNP) cargo (see 1.3.1 Early views on the spatial transcription organization), the transcription factory hypothesis proposes that polymerase machineries are too large to relocate and therefore remain stationary while the template DNA is passed through them.

Transcription factories were reported as clusters of 8-30 RNAPIIs anchored to a nuclear matrix (Martin and Pombo, 2003). Upon activation, genes are brought to these factories and reeled through one or several RNAPIIs. Consequently, the aggregating nascent RNAs (nRNAs) are confined to a small volume, forming a single observable focus (Papantonis and Cook, 2013). The number of transcription foci per

cell can vary depending on the cell type and differentiation state. The clustering of genes in transcription factories was considered as a factor in facilitating the rapid and coordinated expression of co-regulated genes in response to cellular stimuli or during differentiation (Osborne et al., 2004).

1.3.3 The "splicing" model

Splicing, a fundamental process in eukaryotic gene expression, plays a pivotal role in shaping both transcriptional dynamics and genome structure. By removing intronic sequences from pre-mRNA transcripts and ligating exonic sequences together, splicing generates mature mRNA transcripts (Gilbert, 1978). This process greatly expands the coding potential of the genome by allowing the same gene to produce multiple protein isoforms with distinct functions. Additionally, alternative splicing events can modulate gene expression levels and contribute to cellular diversity (Amara et al., 1982; Wang et al., 2008). The splicing process is heavily influenced by chromatin structure and the spatial organization of genes within the nucleus (Tammer et al., 2022).

Nuclear speckles are dynamic membraneless bodies located within the cell nucleus, enriched with RNA and proteins, many of which are involved in the splicing process (Spector and Lamond, 2011). It has been demonstrated that nuclear speckles can increase the transcription of associated genes, likely due to the high concentration of splicing factors in these bodies. Despite lacking direct DNA content, nuclear speckles are closely linked with specific genomic regions and RNAs located at the periphery of the speckles. (Hall et al., 2006; Shopland et al., 2003; Spector and Lamond, 2011). It was noted that some genes and their nascent RNAs often localize with or near nuclear speckles (Jackson et al., 1993; Wansink et al., 1993), and that this co-localization is particularly pronounced for highly transcribed genes with notable examples including globin genes, *SLC4A1*, and *ERAF* (Brown et al., 2008; Hu et al., 2009; Spector, 1993).

1.3.4 The modern knowledge about transcription regulation

The precise regulation of mammalian transcription requires an intricate coordination among various factors, such as transcription factors, the histone code, RNAPII C-terminal domain code, DNA modifications, non-coding RNAs and RNA modifications. To ensure the accurate spatio-temporal regulation of gene transcription, an additional epigenetic layer of regulation has evolved, namely regulatory elements of genes. Among the most extensively studied regulatory elements are enhancers which are believed to interact with gene promoters to regulate transcriptional bursting (Chubb et al., 2006; Raj et al., 2006). It is noteworthy that genes primarily involved in differentiation and development rely heavily on their enhancers for regulation, while housekeeping genes are expressed without such elaborate regulatory structures (Dejosez et al., 2023).

Enhancers, which are short nucleotide sequences in the genome, may reside several kilobases up to megabases away from their target promoter and can be found within intra- or intergenic regions, including both introns and exons of a gene (Borsari et al., 2021; Lettice et al., 2003). Due to their diverse locations and features, mapping putative enhancers still presents significant challenges (Schoenfelder and Fraser, 2019). Emerging evidence indicates that the interaction between a promoter and an enhancer is facilitated by the formation of loops which appear to be transient and not stably established but instead rely on extrusion of chromatin (Deng et al., 2012; Karpinska and Oudelaar, 2023; Tolhuis et al., 2002). These looping interactions contribute to establishing the three-dimensional architecture of the genome by bringing linearly distant regulatory sequences into close spatial proximity within topologically associating domains or insulated neighborhoods. Promoter-enhancer looping may also help define the boundaries and insulation of these regulatory domains, preventing “false” interactions between enhancers and promoters across domain boundaries (Gaszner and Felsenfeld, 2006).

1.3.5 Liquid-liquid phase separation and transcription

An additional factor that is speculated to contribute to the precise regulation of transcription is liquid-liquid phase separation within the nucleoplasm. Biomolecules can dynamically assemble into distinct, membrane-less compartments within the nucleus called transcriptional condensates. These transcriptional condensates are separate from the surrounding nucleoplasm (Hyman et al., 2014). The formation of these transcriptional condensates is driven by multivalent, weak interactions between the intrinsically disordered regions of many proteins involved in transcription, such as TFs, the C-terminal domain of RNAPII, splicing factors, etc., which allow the condensates to rapidly associate and disassociate (Boehning et al., 2018; Hnisz et al., 2017; Wei et al., 2020).

Importantly, the phase separation of transcriptional components has been linked to the three-dimensional reorganization of the genome, as the condensates can cluster and compartmentalize distant genomic regions that share common transcriptional regulators (Erdel and Rippe, 2018; Hildebrand and Dekker, 2020). Furthermore, dysregulation of transcriptional condensate formation has been implicated in various disease states, highlighting their importance in normal cellular function (Cai et al., 2021; Elbaum-Garfinkle, 2019).

1.4 Methodological Considerations

The advancement of modern biology is heavily dependent on the progress made in new methodologies and techniques, particularly in genome engineering and microscopy. Significant progress has been made in enhancing the resolution and automation of microscopes, along with the

development of oligo based probes. Both have revolutionized microscopy elevating it to a level comparable to molecular techniques.

1.4.1 Light microscopy resolution

The disparity between the two perspectives on transcription mechanisms described above can largely be attributed to the limited resolution of light microscopy. Even with the use of super-resolution microscopy, the structure of a single gene or locus remains unresolved, typically appearing as an irregularly shaped spot (Mateo et al., 2019; Schermelleh et al., 2010). In contrast, lateral lampbrush loops and polytene puffs are observable even under a phase contrast microscope, thanks to the substantial length of transcription units and their high transcriptional activity (Bjork and Wieslander, 2015; Macgregor, 1993). Consequently, both sufficient length and a high level of expression are crucial for visualizing expressed genes within interphase nuclei. However, the combination of these two characteristics is rarely found in cultured mammalian cells, which are the primary focus of transcriptional studies (Schoenfelder et al., 2010).

1.4.2 Visualization of DNA and RNA by Fluorescence In Situ Hybridization

Fluorescence in situ hybridization (FISH) is a powerful technique used to study chromatin structure and genome folding in cells. FISH involves the hybridization of (in)directly fluorescently labeled nucleic acid probes to specific DNA sequences of interest within the chromatin (Bauman et al., 1980; Langer-Safer et al., 1982). These probes can be designed to target various genomic regions, such as entire chromosomes or their arms, repetitive sequences, single genes or their regulatory elements.

Large genomic DNA fragments cloned into large-insert vectors, such as plasmids, fosmids or bacterial artificial chromosomes (BACs), are still commonly used in FISH experiments for genomic mapping and gene localization (Cremer et al., 2008; Shizuya et al., 1992). However, they are subject to several drawbacks that can affect the accuracy and reliability of FISH results. Firstly, these vectors have a predetermined insert size, typically ranging up to 300 kb, which may exceed the desired region of interest and thereby limits the genomic resolution (Boettiger and Murphy, 2020). Moreover, they may contain repetitive DNA sequences, such as transposable elements or segmental duplications, which need to be blocked in order to avoid non-specific hybridization and false-positive signals in FISH experiments, obscuring the accurate localization of target genomic regions (Sealey et al., 1985). Additionally, generating probes from large-insert vectors involves labor-intensive procedures like screening, clone isolation, and labeling, making the process time-consuming and resource-intensive (Cremer et al., 2008). Lastly, the available libraries may not cover the entire genome or may not be available for certain species of interest, limiting the applicability of BAC-based FISH approaches.

Recent progress in synthetic DNA production and the accessibility of sequenced genomes has led to the adaption of synthetic DNA oligonucleotides (oligos) as a source of FISH probe material (Figure 3A). Oligo-based probes present significant advantages over conventional probes. These probes can be designed for any sequenced genome, meticulously selected to avoid hybridization to repetitive sequences, and be programmed to incorporate stretches of exogenous sequences that serve as scaffold domains for further hybridization steps (Beliveau et al., 2015) (Figure 3B). The establishment of oligo probes as a primary source of FISH probe material has also lead to the advancement of computational tools for oligo probe design, such as iFISH, OligoMiner and PaintSHOP (Beliveau et al., 2018; Gelali et al., 2019; Hershberg et al., 2021).

Recent advances in multiplexed amplification of FISH signals made it possible to achieve high levels of simultaneous amplification and sequential detection with high sampling efficiency and simple workflows. One of them is the Signal Amplification By Exchange Reaction (SABER) (Kishi et al., 2019), which utilizes oligonucleotide-based FISH probes with long, single-stranded DNA concatemers generated by Primer Exchange Reaction (Kishi et al., 2018). Their polymeric structure serves as a hybridization scaffold for localizing numerous fluorescent imager oligonucleotides at the target sites (Figure 3C), resembling the sequences encountered in other branched signal amplification methods (Player et al., 2001). In our use case, this allows for the detection of small DNA sequences and low-abundance RNA transcripts with high precision and enhanced sensitivity compared to conventional FISH techniques. The signal amplification process in SABER-FISH results in an improved signal-to-noise ratio, enabling clearer and more reliable detection of target sequences. The versatile technique can be adapted for use with different fluorescent dyes and DNA probes, allowing for customization based on the specific experimental requirements and target sequences. SABER-FISH can be multiplexed for the simultaneous visualization of multiple target sequences, enabling the investigation of complex genomic interactions and co-localization studies.

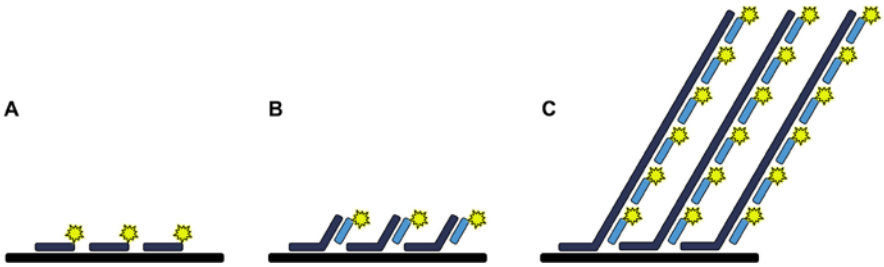


Figure 3: Comparison of Fluorescence In Situ Hybridization Techniques.

(A) Directly labeled FISH Probes. The probes are designed to bind to specific DNA sequences and directly conjugated to a fluorescent dye. **(B)** Indirectly labeled FISH probes use a primary probe that binds to the target DNA, followed by a secondary hybridization with a fluorescently labeled oligo, which hybridizes to the overhang sequence of the primary probes. **(C)** SABER-FISH uses primary probes with long concatemer overhangs that serve as a hybridization scaffold for multiple fluorescently labeled oligos, significantly enhancing signal intensity.

2 Publications

Spatial organization of transcribed eukaryotic genes

Susanne Leidescher, Johannes Ribisel*, **Simon Ullrich***, Yana Feodorova, Erica Hildebrand, Alexandra Galitsyna, Sebastian Bultmann, Stephanie Link, Katharina Thanisch, Christopher Mulholland, Job Dekker, Heinrich Leonhardt, Leonid Mirny and Irina Solovei

Nature Cell Biology, 2022 - https://doi.org/10.1038/s41556-022-00847-6

nature cell biology

ARTICLES

https://doi.org/10.1038/s41556-022-00847-6

Check for updates

Spatial organization of transcribed eukaryotic genes

Susanne Leidescher[1], Johannes Ribisel[2,9], Simon Ullrich[1,9], Yana Feodorova[1,3], Erica Hildebrand[4], Alexandra Galitsyna[5], Sebastian Bultmann[1], Stephanie Link[6], Katharina Thanisch[1,8], Christopher Mulholland[1], Job Dekker[4,7], Heinrich Leonhardt[1], Leonid Mirny[2] ✉ and Irina Solovei[1] ✉

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. Owing to the small sizes of most previously studied genes and the limited resolution of microscopy, the structure and spatial arrangement of a single transcribed gene are still poorly understood. Here we study several long highly expressed genes and demonstrate that they form open-ended transcription loops with polymerases moving along the loops and carrying nascent RNAs. Transcription loops can span across micrometres, resembling lampbrush loops and polytene puffs. The extension and shape of transcription loops suggest their intrinsic stiffness, which we attribute to decoration with multiple voluminous nascent ribonucleoproteins. Our data contradict the model of transcription factories and suggest that although microscopically resolvable transcription loops are specific for long highly expressed genes, the mechanisms underlying their formation could represent a general aspect of eukaryotic transcription.

The understanding of eukaryotic gene transcription and the mechanisms of its regulation is progressively increasing at both the molecular[1–3] and nuclear[4–6] levels. Knowledge pertaining to an intermediate level of transcription organization, that is, the spatial arrangement of a single expressed gene, is, however, surprisingly limited. Early studies on gigantic chromosomes revealed that transcription units form loops emanating from the chromosome axis, so-called lateral loops of lampbrush chromosomes[7] and puffs of polytene chromosomes[8]. It was demonstrated that the 5′ and 3′ ends of loops are fixed in space and RNA polymerases II (RNAPIIs) move along a DNA template carrying a cargo of progressively growing nascent RNA transcripts (nRNAs)[7,8].

More recent studies of interphase nuclei have not revealed loops but identified clusters of expressed genes, aggregations of elongating RNAPIIs and nRNA accumulations[9–11]. These observations led to the popular hypothesis of transcription factories, which postulates that RNAPIIs are immobilized in groups, whereas activated genes of the same or different chromosomes approach a transcription factory and are then reeled through immobilized RNAPIIs extruding nRNAs in a single spot[12,13].

The discrepancy between these two views on transcription mechanisms can be explained to a great extent by the limited resolution of light microscopy. Even when using super-resolution microscopy, the structure of a single gene or locus is not resolved and, as a rule, is represented by an irregularly shaped spot[14–16]. In contrast, lateral lampbrush loops and polytene puffs are visible even under a phase contrast microscope[7,8,17], owing to the substantial length of transcription units (up to hundreds of kilobases) and their high transcriptional level. Therefore, for the visualization of expressed genes within interphase nuclei, both sufficient length and sufficiently high expression are essential. The combination of these two traits, however, is rarely met in cultured mammalian cells, which are the major source of knowledge about transcription. Indeed, the majority of studied highly expressed genes are short[11], and given that a 10-kb-long gene, when fully stretched, measures only 0.5 µm, it is comprehensible why their structure cannot be resolved by conventional microscopy with a maximum possible resolution of 0.2–0.3 µm[18]. At the same time, long genes are generally not highly expressed, especially not in cultured cells[19].

In this Article, to fill the gap in the knowledge about the spatial organization of transcription, we selected several genes that are both long and highly expressed and studied their spatial arrangement in differentiated mouse cells. We demonstrated that these genes form microscopically resolvable transcription loops (TLs) similar to lampbrush loops and polytene puffs. We provided evidence that TLs are decorated by elongating RNAPIIs moving along the gene axis and carrying nRNAs undergoing co-transcriptional splicing. Furthermore, we showed that long highly expressed genes dynamically modify their harbouring loci and extend into the nuclear interior presumably because of their increased stiffness resulting from decoration with bulky nascent ribonucleoproteins (nRNPs). Collectively, our data indicate that although microscopically resolvable TLs are specific for long highly expressed genes, the mechanisms underlying their formation could be universal for eukaryotic transcription.

[1]Department of Biology II, Biozentrum, Ludwig-Maximilians University Munich (LMU), Planegg-Martinsried, Germany. [2]Institute for Medical Engineering and Science, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Medical Biology, Medical University of Plovdiv; Division of Molecular and Regenerative Medicine, Research Institute at Medical University of Plovdiv, Plovdiv, Bulgaria. [4]Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA. [5]Skolkovo Institute of Science and Technology, Skolkovo, Russia. [6]BioMedizinisches Center, Ludwig-Maximilians University Munich, Planegg-Martinsried, Germany. [7]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [8]Present address: Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany. [9]These authors contributed equally: Johannes Ribisel, Simon Ullrich. ✉e-mail: leonid@mit.edu; Irina.Solovei@lrz.uni-muenchen.de

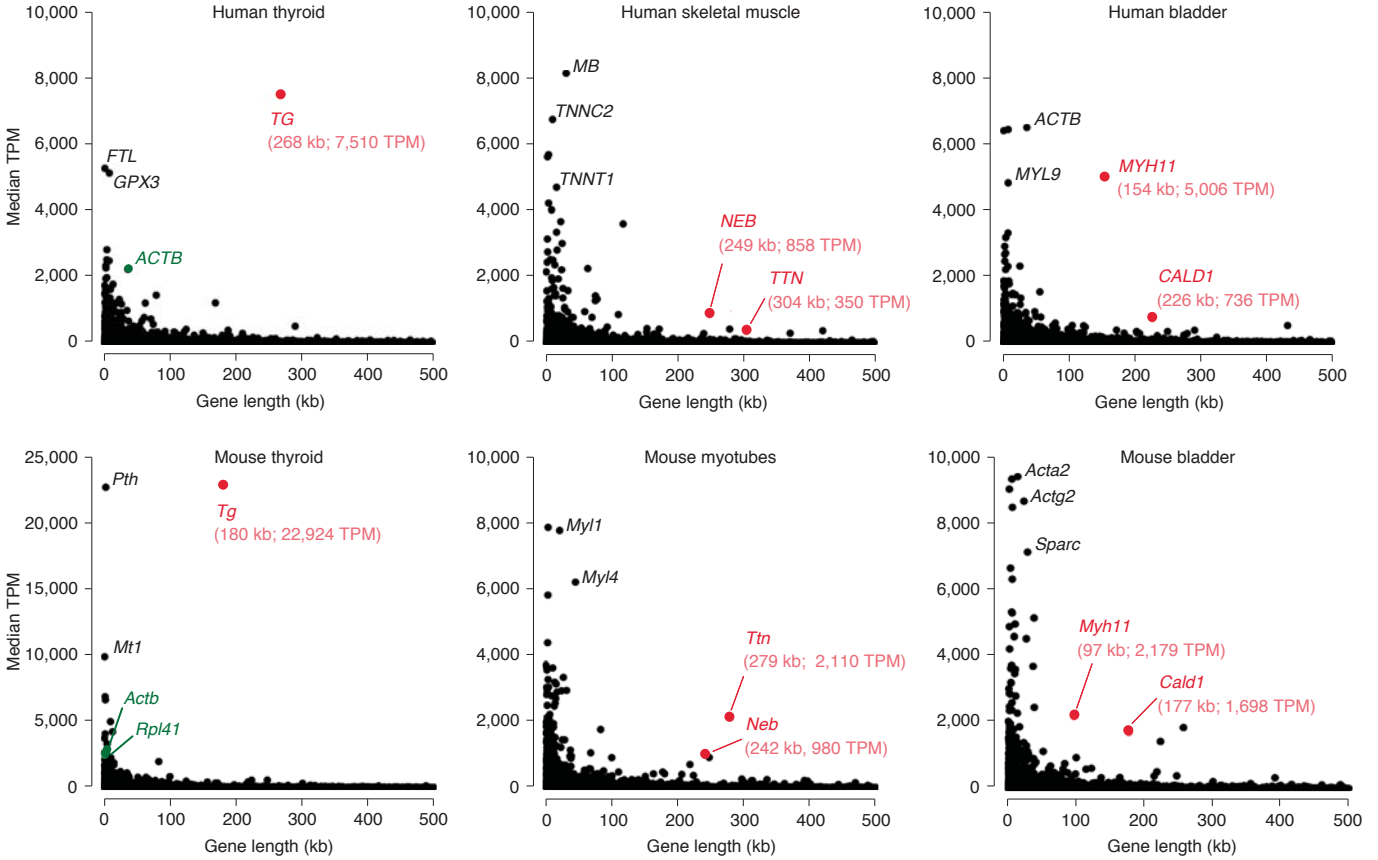


Fig. 1 | Selection of long highly expressed genes. Top: analysis of gene expression in selected human tissues (retrieved from the GTEx database). Bottom: RNA-seq analysis of corresponding mouse tissues and cells from this study. Expression level (median TPM) is plotted against gene length according to 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 the expression of housekeeping genes, such as *Actb* and *Rpl41* (marked in green). For data on all protein coding genes, see Supplementary Tables 1 (human) and 2 (mouse).

Results

Selection of highly expressed long genes. For microscopic visualization, we searched for genes that are both relatively long and highly expressed, using thresholds for length of ≥100 kb, corresponding to the size of the smallest discernible lampbrush loops[20], and for expression level of ≥1,000 transcripts per million (TPM), corresponding to the average expression level of the human *GAPDH* gene (Genotype-Tissue Expression (GTEx) Consortium).

The analysis of gene lengths and expression levels within the human and mouse genomes showed that less than 20% of the genes are ≥100 kb (Extended Data Fig. 1a) and that such genes, as a rule, are lowly expressed (Extended Data Fig. 1b–d). On the basis of GTEx RNA sequencing (RNA-seq) data, we selected ten human genes that were above our thresholds and out of these further selected those that are expressed in cell types unambiguously identifiable in tissue sections (Supplementary Table 1). The most highly expressed gene among them was the thyroglobulin gene (*TG*, 268 kb, 7,510 TPM) coding for the extracellular protein thyroglobulin secreted by thyrocytes; the other four selected genes encode structural proteins of the contractile machineries of skeletal muscle (titin, *TTN*; nebulin, *NEB*) and smooth muscle (caldesmon 1, *CALD1*; myosin heavy chain 11, *MYH11*) (Fig. 1).

Next, we performed RNA-seq analysis of mouse thyroid gland, myotubes, myoblasts and bladder to confirm the high expression of these five genes in the corresponding mouse cells (Supplementary Table 2). In particular, the mouse thyroglobulin gene (*Tg*, 180 kb) is exceptionally upregulated (22,924 TPM, constituting 2.5% of all sequence reads) with an expression level ten times higher than that of some ubiquitously expressed housekeeping genes, such as non-muscle β-actin (*Actb*, 2,791 TPM) or large ribosomal subunit protein EL41 (*Rpl41*, 2,467 TPM). All five selected mouse genes—*Tg*, *Ttn*, *Neb*, *Cald1* and *Myh11*—met the conditions we have defined as prerequisites for our study (Fig. 1).

Highly expressed long genes form TLs. Using genomic probes (bacterial artificial chromosomes, BACs) encompassing the selected genes (Supplementary Table 3), we carried out DNA fluorescence in situ hybridization (FISH) on cryosections of the corresponding mouse tissues—thyroid gland (*Tg*), skeletal muscle (*Ttn* and *Neb*), heart muscle (*Ttn*), bladder and colon (*Myh11* and *Cald1*)—as well as on cultured myoblasts (*Cald1*) and myotubes (*Ttn* and *Neb*). 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 genes. In the non-expressed state, genes were condensed into single compact foci sequestered to the nuclear periphery (Fig. 2a). In the expressed state, genes were strongly decondensed and exhibited several smaller foci in the nuclear interior (Fig. 2b and Supplementary Fig. 1).

Next, we performed RNA FISH on the same set of samples, omitting both RNase treatment and cellular DNA denaturation. With this setup, the same genomic probes used for DNA FISH hybridize to nRNAs. As expected, cells not expressing the genes were lacking FISH signals, whereas expressing cells exhibited massive RNA signals (Fig. 2c, Extended Data Fig. 2 and Supplementary Fig. 1).

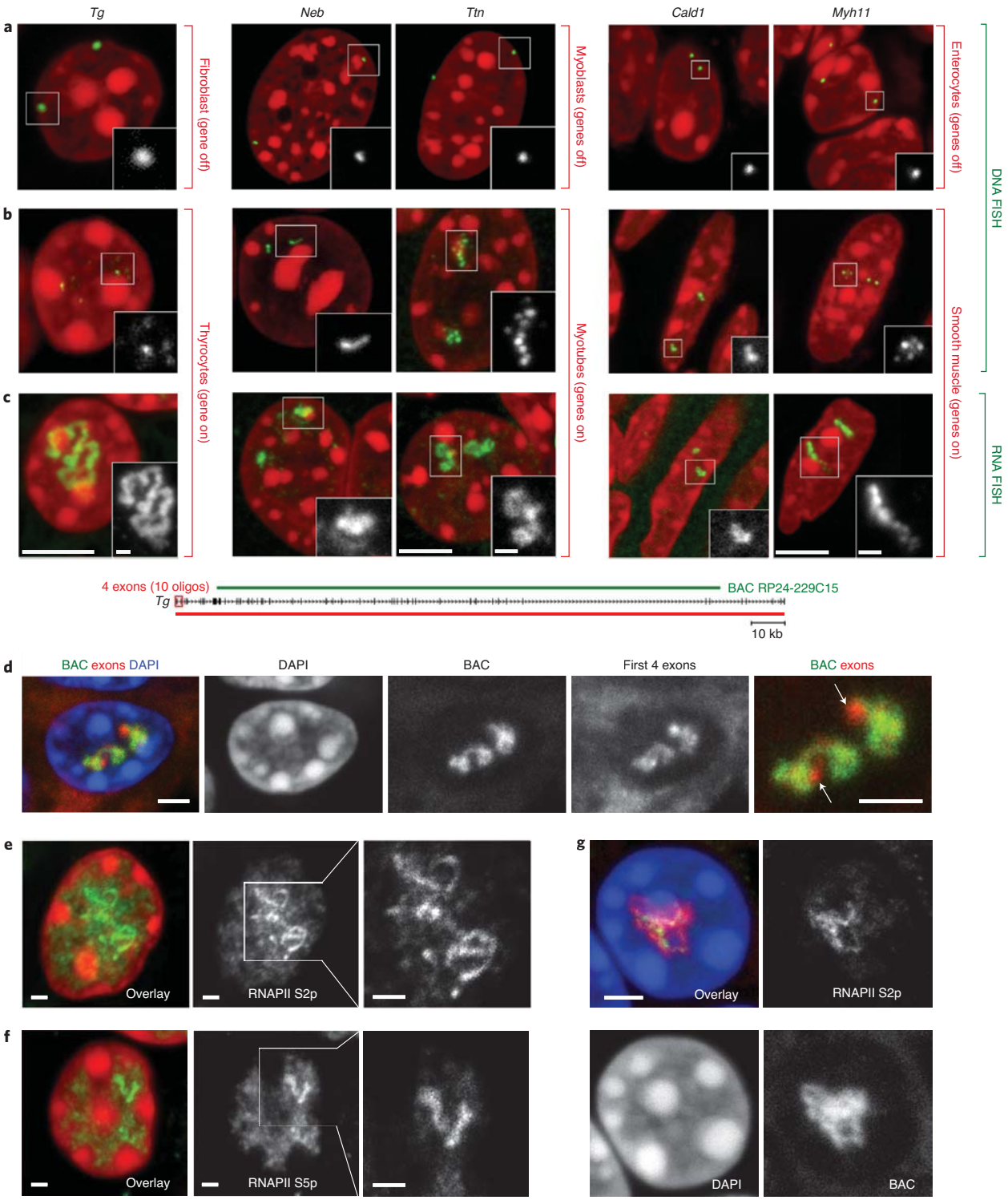


Fig. 2 | Highly expressed long genes form TLs. a–c, Five selected genes after either DNA FISH (**a**,**b**) or RNA FISH (**c**) with corresponding genomic probes. Control cells not expressing the respective genes exhibit focus-like condensed DNA signals at the nuclear periphery (**a**); in expressing cells, genes are strongly decondensed (**b**). RNA FISH reveals TLs by hybridization to multiple nRNAs decorating the genes (**c**). **d**, The entire *Tg* TL detected by an oligoprobe, hybridizing to the first 5′ exons. The schematic shows the distribution of the BAC covering the mid-part of *Tg* (green line) and the oligoprobe for the first four exons (red rectangle) *labeling the entire Tg* gene (red line). Arrows point at *Tg* TL regions labelled by the oligoprobe but not by the BAC probe. **e**,**f**, *Tg* TLs revealed by immunostaining of the elongating (**e**, Ser2p) and initiating (**f**, Ser5p) forms of RNAPII. **g**, ImmunoFISH showing co-localization of structures marked with elongating RNAPII and *Tg* TLs. Images are projections of 1–2.5 µm confocal stacks. Scale bars: 5 µm, in insets, 1 µm (**a**–**c**); 2 µm (**d**,**g**); 1 µm (**e**,**f**). Data represent ten (**a**–**c**) and two (**d**–**g**) independent experiments.

The signals were extended and either had the shape of a coiled loop (for example, *Tg* and *Ttn*), or formed less discernible elongated structures (for example, *Neb* and *Cald1*). The loops were particularly prominent in the case of the *Tg* gene—they were spread throughout the nuclear interior and measured up to 10 µm (Extended Data Fig. 2a, Supplementary Fig. 1 and Supplementary Video 1).

Since introns, as a rule, are substantially longer than exons, we assumed that the used genomic probes hybridized mostly to unspliced introns of nRNAs, thereby outlining the contours of transcribed genes. An expressed gene can also be visualized in its entirety by using oligoprobes hybridizing to the first 5′ exons (Fig. 2d). In agreement with FISH data, immunostaining of initiating RNAPII (Ser5p) in thyrocytes highlighted short stretches of the *Tg* gene, whereas immunostaining of elongating RNAPII (Ser2p) revealed strongly convoluted structures (Fig. 2e,f) co-localizing with RNA FISH signals (Fig. 2g) and apparently corresponding to the *Tg* gene axis covered by polymerases. Hereafter we refer to these structures as transcription loops (TLs).

TL signals varied in shape and degree of coiling, and no consistent pattern of loop folding or position was observed with the exception of their invariably interior nuclear location. Importantly, almost every single fully differentiated thyrocyte, skeletal muscle, heart muscle or smooth muscle cell exhibited two TLs corresponding to two alleles. Scoring of RNA FISH signals revealed that only a very small proportion of nuclei had one TL, as a result of either a long transcription pause or monoallelic expression (Supplementary Table 4 and Supplementary Fig. 2).

At the level of light microscopy, one can trace only the general contour of the TLs, but within this contour there is a finer coiling of the gene axis not resolvable by deconvolution or high-resolution microscopy (Extended Data Fig. 3a). To demonstrate the inner structure of TLs, we performed FISH with the *Tg* probe on thin thyroid sections (50–70 nm) and indeed observed such coiling (Extended Data Fig. 3b). Therefore, the compaction level of *Tg* TLs compatible with nucleosomal chromatin (ca. 17 kb µm^{-1}), which was calculated based on contour length measurements (Extended Data Fig. 3c), is an overestimation. We cannot exclude that during *Tg* transcription a substantial proportion of nucleosomes is lost, in agreement with the observations made for highly expressed ribosomal genes[21] or lampbrush chromosome lateral loops[22].

TLs manifest progression of transcription and splicing. To confirm that the observed RNA signals are not accumulations of messenger RNAs but represent nRNAs, we performed two types of RNA FISH experiment. First, we used differentially labelled probes hybridizing to 5′ exons (2–12) and 3′ exons (33–47) of the *Tg* gene. We reasoned that if loops represent mere mRNA accumulations, both probes would label the entire structures. The probe for 3′ exons, however, labelled only the second half of TLs (Fig. 3a), confirming that TLs are decorated with growing nRNAs. On a smaller scale, we demonstrated that the oligoprobe for the 5′ half of *Tg* long intron 41 labels the entire intron, whereas the oligoprobe for the 3′ half of the intron labels only half of it (Fig. 3b).

Second, we used genomic probes highlighting 5′ and 3′ introns. We expected that the two probes hybridize to nRNAs on TLs in a consecutive manner as a result of co-transcriptional splicing and, indeed, observed such a pattern. The 5′ genomic probe strongly labelled the first half of the loop by hybridizing to nRNAs containing yet unspliced 5′ introns and the 3′ probe labelled the second half of the loop by hybridizing to yet unspliced 3′ introns (Fig. 3c,d). The oligoprobes covering 1–5 kb of two sequentially positioned introns of *Tg*, *Ttn* and *Cald1* labelled TLs only partially, producing separate and non-overlapping signals (Extended Data Fig. 4), further confirming the ongoing splicing[23]. In agreement with multiple splicing events occurring along highly expressed genes, TLs are decorated by components of exon-junction complexes and are either closely adjacent to, or partly co-localize with, nuclear speckles (Fig. 3e and Extended Data Fig. 5).

Besides TLs, RNA FISH with genomic probes revealed numerous granular signals scattered throughout the nucleoplasm, representing either mRNA or accumulations of spliced-out but not yet degraded introns (for detailed description, see Extended Data Fig. 6).

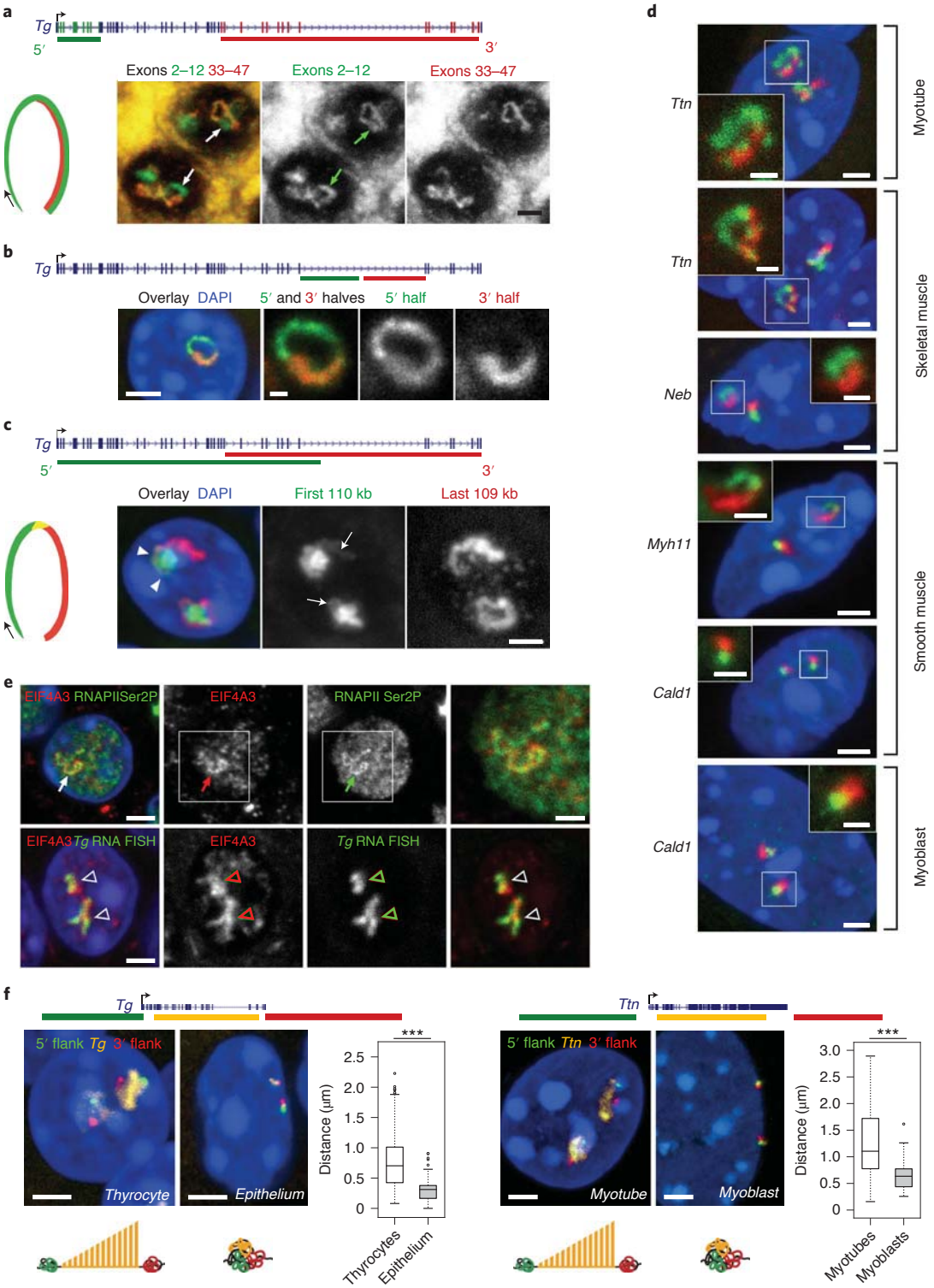
TLs are open loops with separated flanks. One of the popular views on the structure of transcribed genes is that the physical interaction between the transcription start site (TSS) and transcription termination site (TTS) is crucial for a high level of gene expression. Such an interaction has been proposed to act as a 'bridge' for RNAPIIs enabling them to immediately re-initiate transcription after termination, that is, to transcribe 'in circles'[24]. To assess whether the beginnings and the ends of the selected genes are proximal, we visualized TLs and their genomic flanks for *Tg*, *Ttn* and *Myh11* genes and found that the 5′ and 3′ gene flanks are visibly 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 are around two times larger compared with those in control cells not expressing the genes (Fig. 3f). Together these findings demonstrate that, contrary to the 'transcription cycle' hypothesis, TLs of long highly expressed genes are open-ended loops with separated flanks.

TLs are larger for more highly transcribed genes. Although the different hybridization efficiencies of the used probes and the various image acquisition conditions do not allow a legitimate quantitative comparison of DNA and RNA signal sizes between genes, the qualitative comparison of signals suggests that gene length alone is not a good predictor of TL size. The TL size is rather defined

Fig. 3 | TLs manifest transcription progression and dynamically modify the harbouring chromosomal loci. a, Successive labelling of *Tg* TLs with probes for 5′ (green) and 3′ (red) exons reflecting changes in exon composition. Arrows point at the TL regions labelled with only the 5′ probe. **b**, Successive labelling of the *Tg* long intron 41 with oligoprobes for the entire intron (green) and for its second 3′ half (red). **c**, Sequential labelling of *Tg* TLs with genomic probes highlighting 5′ (green) and 3′ (red) introns reflecting changes in the intron composition. The 5′ genomic probe includes 5′ exons and thus faintly labels the second half of the loop (arrows). Arrowheads mark the region hybridized by both overlapping probes. **d**, TLs formed by other long highly expressed genes exhibit co-transcriptional splicing. **e**, Co-localization of the component of the exon-junction complex EIF4A3 (red) with *Tg* TLs (green) visualized either by RNAPII staining or by RNA FISH. In **a**–**e**, data represent three independent experiments. **f**, Expressed genes expand and separate their flanks. Distances between 5′ (green) and 3′ (red) flanking regions of the *Tg* and *Ttn* genes (yellow) are larger in cells expressing the genes compared with control cells with silent genes. Projections of confocal sections through 1–3 µm. Scale bars: 2 µm; in insets, 1 µm (**d**,**e**). The distribution of the used probes with respect to the studied genes are depicted above the image panels. The boxplots depict the 3D distance between the flanking regions in expressing and non-expressing cells. The median inter-flank distance for *Tg* in thyrocytes is 2.3-fold larger than in control epithelial cells (703 nm versus 311 nm) ($n = 203$ alleles in thyrocytes and 180 alleles in epithelial cells across two independent experiments, $^{***}P = 2.167 \times 10^{-34}$, two-sided Wilcoxon rank sum test). The median inter-flank distance for *Ttn* in myotubes is 1.7-fold larger than in control myoblasts (1,104 nm versus 634 nm) ($n = 117$ alleles in myotubes and 63 alleles in myoblasts across three independent experiments), $^{***}P = 2.584 \times 10^{-11}$, two-sided Wilcoxon rank sum test). Boxplots show the median (horizontal line), 25th to 75th percentiles (boxes) and 90th percentile (whiskers) of the group.

by expression level: for example, RNA FISH signals of the longer *Ttn* gene are less expanded compared with those of the shorter but ten times more highly expressed *Tg* gene (Fig. 4a,b). Not surprisingly, we also noticed that the size of RNA FISH signals inversely correlates with the condensation of genes detected by DNA FISH: the gene body of *Tg* is strongly decondensed and marked with only few faintly labelled foci, whereas the *Ttn* gene is more condensed and displays a chain of several larger foci (Fig. 4a,b). The *Neb* gene, expressed at a level that is more than two times lower than that of *Ttn*, typically exhibits even more prominent gene body condensation (Fig. 4c). The body of the dystrophin gene (*Dmd*), the longest mammalian gene (ca. 2.3 Mb) with very low expression, remains fairly condensed (Fig. 4d).

The coordinated pattern of gene body condensation and TL expansion presumably reflects the frequency and duration of transcriptional bursts that are directly related to the expression level of genes[25]. A possible scenario is that a highly expressed gene is transcribed in long and frequent bursts; an intermediately expressed



gene in short, less frequent bursts; and the bursts of a lowly expressed gene are even shorter and less regular (Fig. 4). Consequently, in the case of high expression, a gene is covered by long and frequent RNAPII convoys[26] and the entire gene body forms a TL (for example, *Tg* and *Ttn*). In the case of lower expression, short and infrequent bursts do not decondense the entire gene but rather form short and sparse RNPII convoys leading to small 'travelling' TL (for example, *Neb* and *Cald1*). In both cases, however, transcription occurs via the same mechanism with RNAPIIs moving along genes, which is evident from the sequential highlighting of introns (Fig. 3c,d).

TLs are dynamic structures. To confirm that TLs are 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 nuclease-deactivated Cas9 (dCas9) conjugated with the tripartite transcription activator VP64-p65-Rta (VPR)[27]. Targeting the dCas9-coupled activator to the *Ttn* promoter region induced expression of the full-length gene in myoblasts (Supplementary Fig. 3) to a level comparable to that in myotubes, which was accompanied by formation of *Ttn* TLs and by a twofold increase of *Ttn* median inter-flank distances (Fig. 5a).

Conversely, we aimed to eliminate *Ttn* TLs in differentiated myotubes by transcription inhibition. Incubation of myotubes with α-amanitin or actinomycin D led to abortion of RNAPIIs and disappearance of TLs (Fig. 5b). In agreement with their different mechanisms of action[28], the drugs affected gene body condensation differently. In the case of α-amanitin, which prevents DNA and RNA translocation by binding near the catalytic site of RNAPII causing its stalling and degradation, the *Ttn* gene condensed and the gene flanks converged. In the case of actinomycin D, which prevents transcription initiation and RNAPII progression by intercalating into the DNA minor groove, *Ttn* remained decondensed with separated flanks (Fig. 5b).

To further monitor TL dynamics, we treated differentiated myotubes with 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), a drug reversibly preventing transcription elongation[28]. As anticipated, during the first 2 h of drug treatment, nRNA signals diminished and the gene body condensed. After drug removal, the TL signal gradually recovered with the 5′ signal emerging first and the 3′ signal emerging only after ca. 1 h (Fig. 5c). In accordance with transcriptional changes, *Ttn* flanks converged upon drug application and diverged after drug removal (Fig. 5d).

TLs can expand from harbouring chromosomes. Although the internal position of genes within their chromosome territory is compatible with transcription[29], very gene-dense and highly transcriptionally active chromosomal regions have been shown to significantly extend from chromosome territories for reasons that are not quite understood[30–34]. To investigate the relationship between TLs and their harbouring chromosomes, we visualized the *Tg* TL and chromosome 15 by FISH detecting both DNA and RNA.

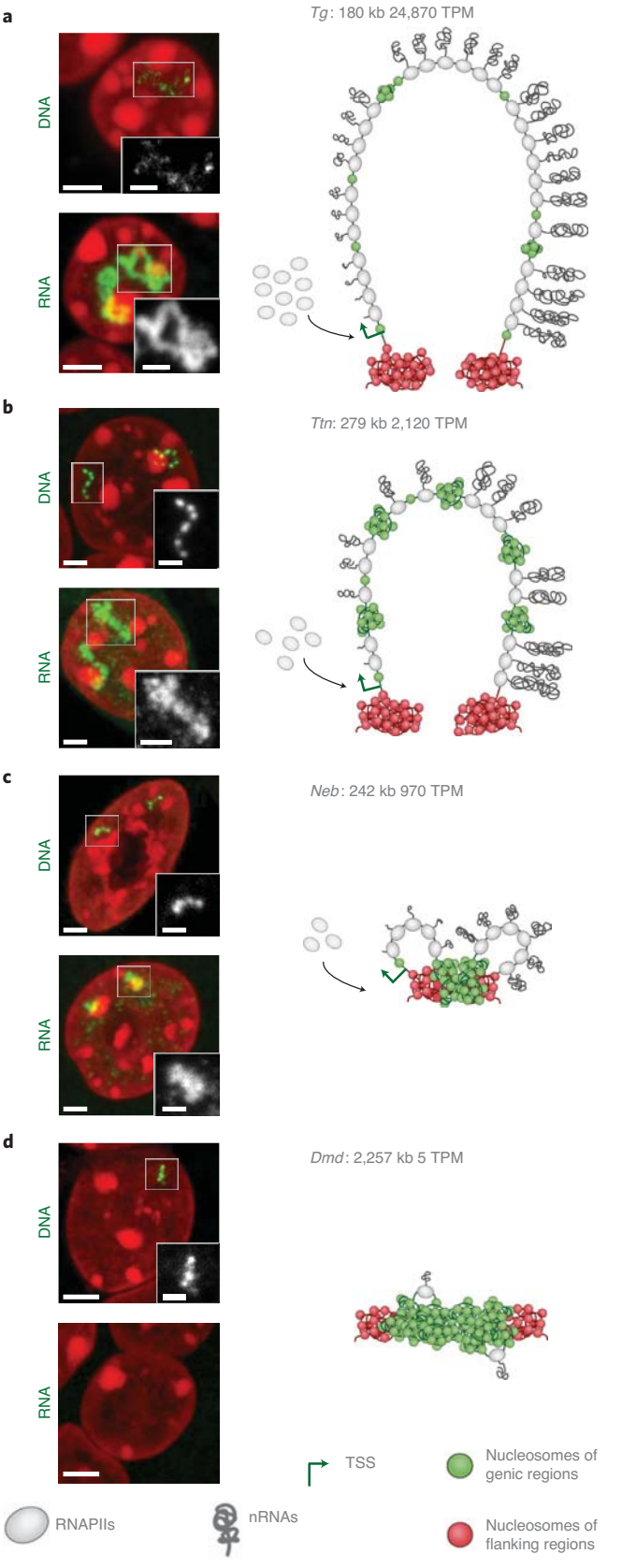


Fig. 4 | TL size is not determined by gene length but by expression level. a–d, Examples of four long genes—*Tg* (**a**), *Ttn* (**b**), *Neb* (**c**) and *Dmd* (**d**)—arranged from top to bottom according to their expression level in the respective cell type. Gene length and expression level are indicated next to the gene symbol. For every gene, two images are displayed on the left, showing DNA FISH detecting the gene body and RNA FISH detecting nRNAs. The more expressed a gene is, the less solid the DNA signal is and the more expanded the RNA signal is. Vice versa, the less expressed a gene is, the more condensed the gene body is and the less extended the RNA signal is. The schematics on the right are speculative interpretations of the observed FISH signal patterns in terms of transcriptional bursts, depicted as RNAPII convoys with attached nRNAs, and transcriptional pauses, depicted as condensed chromatin (green nucleosomes). For simplicity, splicing events are not shown on the schemes. Microscopic images are projections of 1–1.5 µm confocal stacks. Scale bars: 2 µm; in insets, 1 µm. Data represent ten (**a–c**) and four (**d**) independent experiments.

We found that *Tg* TLs are excluded from chromosomes and either expand into the nucleoplasm or coil next to the chromosome territory, forming their own 'transcription territories' (Fig. 6a). The areas of TLs are often depleted of chromatin, indicating an extremely high concentration of proteins involved in the massive transcription of *Tg* (Fig. 6b). Remarkably, in 2% of *Tg* alleles, TLs split chromosome 15 territories into two halves with a gap marked by the 5′ and 3′ *Tg* flanking sequences and filled with *Tg* TLs (Fig. 6c,d).

To further study the arrangement of TLs within their harbouring chromosomes, we performed Hi-C analysis of mouse thyroid, bladder, cultured myoblasts and myoblast-derived differentiated myotubes (Fig. 6e, Extended Data Fig. 7 and Supplementary Fig. 4). We reasoned that if an expressed gene loops out of the harbouring chromosome, its *cis* contacts are diminished compared with the *cis* contacts of the same gene in a silent state, that is, the *cis*-to-*trans* contact frequency ratio is expected to be low, indicating externalization of the region. Analysis of Hi-C data revealed that *cis*-to-*trans* tracks of genes exhibiting the largest TLs (*Tg*, *Ttn*, *Myh11* and *Cald1*, but not *Neb*, which has lower expression) have dips, that is, the genes are externalized, in contrast to the same genes in the silent state or other long but weakly expressed genes (Fig. 6f,g). Notably, upon activation, *Tg*, *Ttn*, *Neb* and *Myh11* relocate from B to A compartment (Extended Data Fig. 8a), while *Cald1* remains in A. While A compartment regions are known to be externalized from their harbouring chromosomes[35–37] (Extended Data Fig. 8b), most TL-forming genes stand out in their degree of externalization compared with other genes on the same chromosome and with the same compartment signal (Extended Data Fig. 8d).

We asked whether similar mechanisms can lead to externalization of other active genes. While lower *cis*-to-*trans* ratio is associated with higher transcription and increasing gene lengths, these trends are due to a few (~100) very highly transcribed or very long genes (Extended Data Fig. 8d,e). Among highly transcribed (TPM > 1,000) genes, longer genes are more externalized ($R^2 = 0.2$–0.5; Fig. 8e). Conversely, for long genes (>50 kb), the externalization increases with transcription ($R^2 = 0.12$–0.28 for TPM = 10^2–10^4). For all other genes, the correlation of externalization with transcription and length is minute ($R^2 < 10^{-3}$). Thus, the externalization of

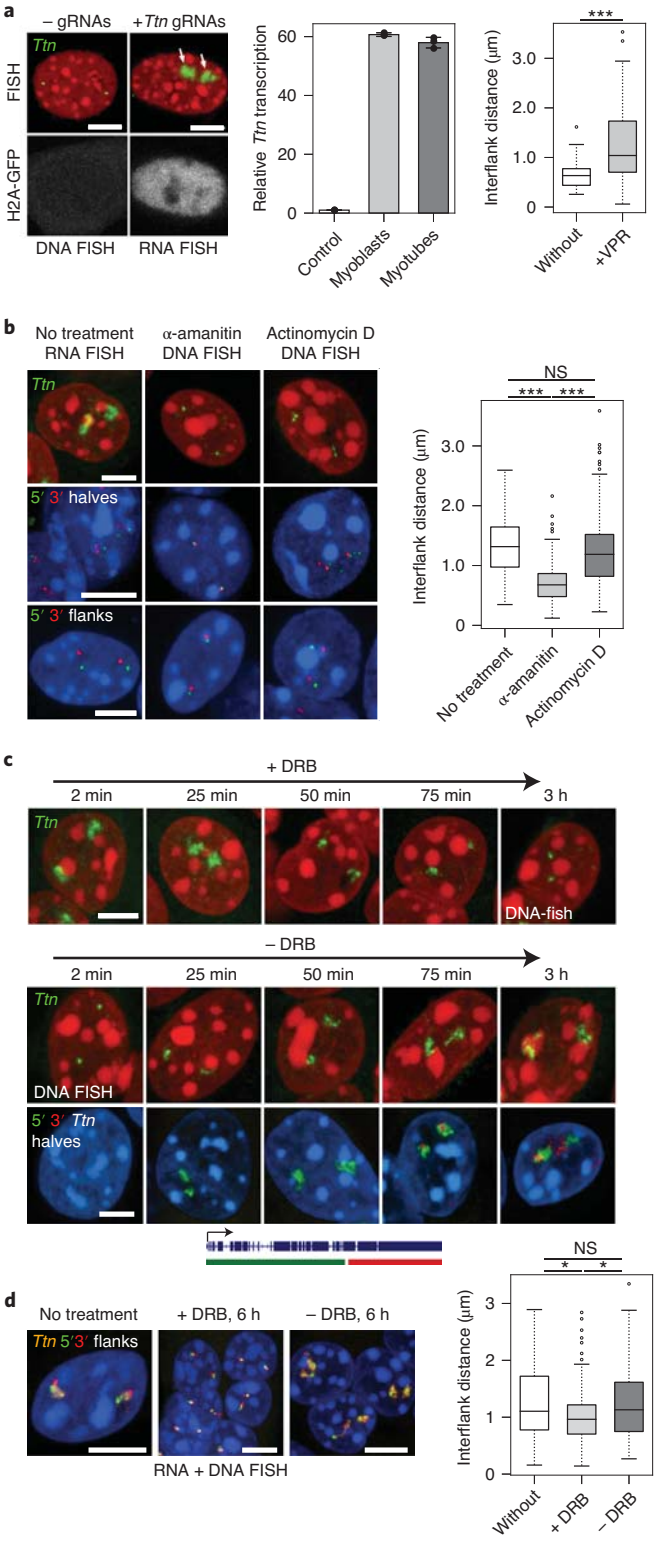


Fig. 5 | TLs are dynamic structures. a, Ectopic induction of *Ttn* TLs in myoblasts by transfection of dCas9-VPR-expressing myoblasts with gRNAs targeting the *Ttn* promoter combined with H2A-GFP-plasmid. In this way, *Ttn* expression was induced to the level of myotubes ($n = 3$, error bars depict standard deviations). Ninety percent of transfected myoblasts exhibit *Ttn* TLs and a twofold increase in median inter-flank *Ttn* distances (1,039 nm versus 634 nm) ($n = 104$ and 63 alleles collected from 52 and 70 cells, respectively, across two independent experiments), $^{***}P = 1.052 \times 10^{-7}$, two-sided Wilcoxon rank sum test. **b**, The transcription inhibitors α-amanitin and actinomycin D eliminate *Ttn* TLs in myotubes. α-amanitin treatment causes strong condensation of the gene body and convergence of the flanks, whereas actinomycin D leaves the gene body decondensed and flanks diverged ($n = 62$, 79 and 112 alleles without treatment, α-amanitin treatment and actinomycin D treatment, respectively, collected from 140 myotube nuclei across three independent experiments), $^{***}P = 5.5 \times 10^{-12}$, not significant (NS) $P = 0.110$ for untreated cells versus α-amanitin treatment and actinomycin D treatment, respectively, $^{***}P = 4.669 \times 10^{-10}$ for α-amanitin treatment versus actinomycin D treatment, two-sided Wilcoxon rank sum test. **c**, Top: DRB treatment leads to a gradual *Ttn* TL shrinkage. Middle: removal of DRB leads to a gradual TL restoration. Bottom: differential labelling of the two halves of *Ttn* (see schematics below the panel) allows to follow emergence first of the 5′ (green) and then, after 1 h, the 3′ (red) signals. **d**, *Ttn* inter-flank distances decrease after DRB treatment (1,104 nm to 963 nm) but remain larger than inter-flank distances in myoblasts (634 nm); upon transcription restoration, the distances are restored (1,130 nm) ($n = 117$, 114 and 88 alleles collected from 160 myotube nuclei without treatment, after 6 h of DRB treatment and 6 h after DRB removal, respectively, across two independent experiments), $^{*}P = 0.013$, NS $P = 0.899$, $^{*}P = 0.026$ for DRB-treated cells versus non-treated cells, cells after DRB removal versus non-treated cells and DRB-treated cells versus cells after DRB removal, two-sided Wilcoxon rank sum test. Scale bars, 5 μm. **a**, **b** and **d** boxplots show the median (horizontal line), 25th to 75th percentiles (boxes) and 90th percentile (whiskers) of the group.

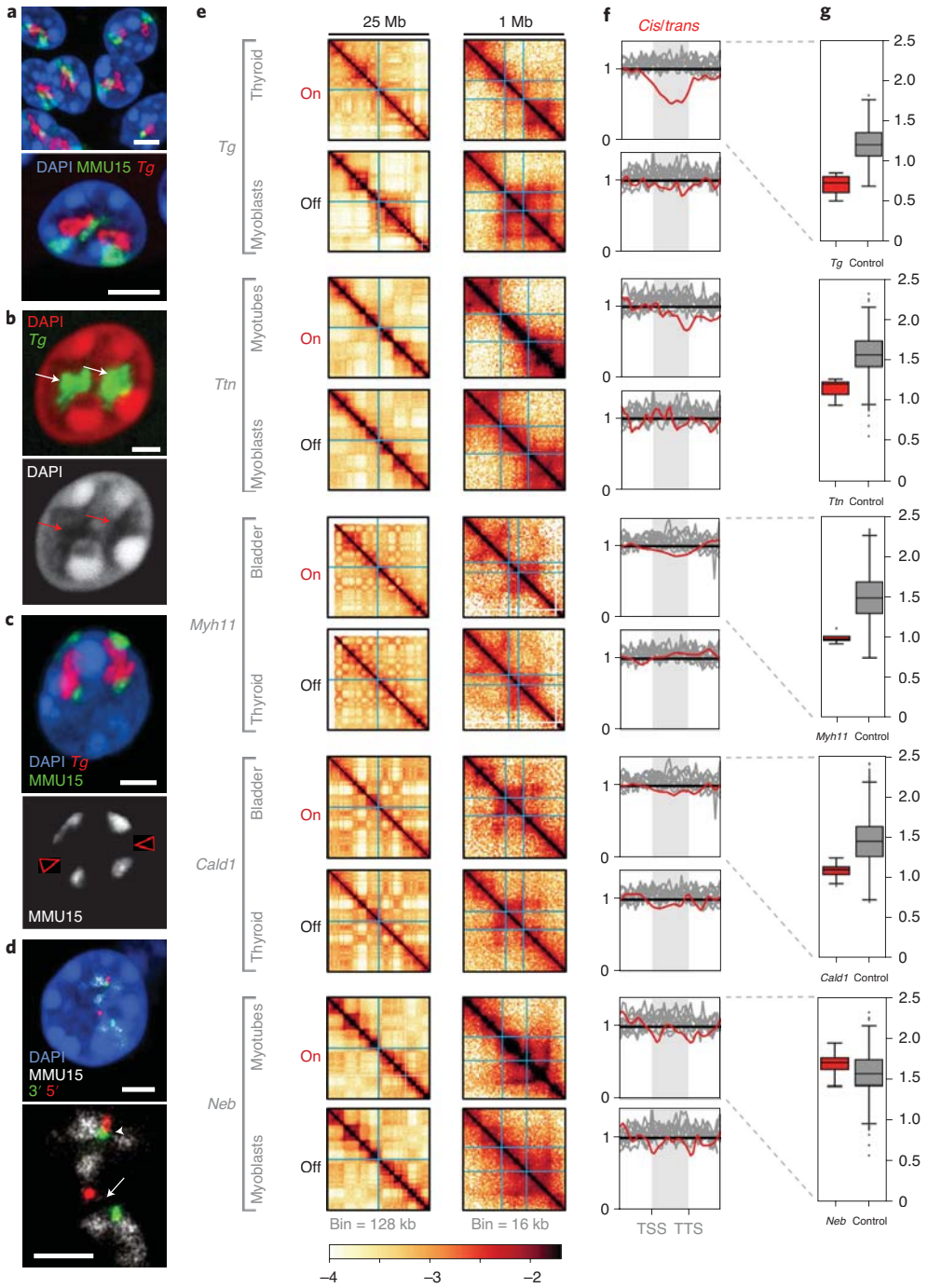


Fig. 6 | TLs are excluded from harbouring chromosomes. a, *Tg* TLs (red) emanate from the harbouring chromosome 15 (green) and protrude into the nuclear interior. **b**, Nucleoplasmic regions occupied by *Tg* TLs (white arrows) are depleted of DAPI-stained chromatin (red arrows). **c**, Chromosomes 15 (green) are split into two halves by *Tg* TLs (red) with unpainted gaps (arrowheads). **d**, The gap in the chromosome territory (arrow) is marked with the 5′ (red) and 3′ (green) flanking sequences. The other chromosome territory is not split (arrowhead). Projections of confocal stacks through 1 µm; scale bars: 5 µm (**a**), 2 µm (**b**–**d**). In **a**–**d**, data represent four independent experiments. **e**, A 25 Mb Hi-C contact map and a 1 Mb zoom view of five genes in active (on), or silent (off) states. TSS and TTS of the genes are marked with light-blue lines (for detailed maps, see Extended Data Fig. 7) The scale at the bottom indicates Log_{10} of contact frequency. **f**, *Cis*-to-*trans* ratio profiles, that is, 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 for the gene of interest (red) and compared with ten other long but lowly expressed genes (grey, see Supplementary Table 2 for the lists of genes). For comparison of *cis*-to-*trans* ratio profiles, the *x* coordinates are rescaled so that the TSSs and TTSs of all genes are aligned (shaded areas). To highlight potential dips localized in the gene body against longer range variations, *cis*-to-*trans* ratio profiles are normalized to unity in the region outside the gene body. **g**, Statistical analysis of *cis*-to-*trans* ratios at a bin size of 32 kb for five studied genes (red) versus long but weakly expressed genes (grey). *Tg*, *Ttn*, *Myh11* and *Cald1* have significantly lower *cis*-to-*trans* ratios compared with control genes (0.735 versus 1.196; 1.213 versus 1.563; 0.987 versus 1.498; 1.102 versus 1.453, respectively) ($P < 0.01$, pairwise Mann–Whitney rank test for smaller values in the target group). Boxplots show the median (horizontal line), 25th to 75th percentiles (boxes) and interquartile range extended by 1.5 (whiskers) of the group. Hi-C assay was performed in three replicates.

genes from their harbouring chromosomes is specific for long and highly expressed genes.

In contrast to the large-scale reorganization of TLs seen in microscopy, their effects in Hi-C are local as they do not alter long-range contacts across chromosomes (Extended Data Fig. 9). We noticed two effects of TLs in Hi-C: (1) they perturb local cohesin-mediated organization (*Tg*, *Ttn*, *Neb* and *Myh11*) largely by diminishing or re-organizing TAD borders (*Ttn*, *Neb* and *Myh11*) and 'dots' (*Tg* and *Myh11*), and (2) several TLs show enrichment of self-interactions visible as 'necks' (*Ttn* and *Neb*) or micro-domains (*Myh11*) (Extended Data Fig. 7), similar to the observations in ref. [38]. Both phenomena could be related to the interplay of transcription and loop extrusion[39–41]. We concluded that TLs perturb the organization of their neighbourhoods, leaving most of the harbouring chromosomes unaffected.

TLs are stiff structures. What makes highly expressed genes expand from chromosomes and separate their flanks? Every RNAPII is associated with an nRNP complex formed by newly synthesized nRNA and numerous proteins involved in RNA processing, quality control, transport and translation[42] (Fig. 7a, left). Therefore, nRNPs containing long nRNAs are bulky structures[8] exceeding the size of nucleosomes (10 nm)[43] or RNAPIIs (14 nm)[44]. Since, during a transcription burst, RNAPIIs travel along a DNA template as a tightly spaced convoy[26], we hypothesized that the dense decoration with bulky nRNPs turns a highly expressed gene into a stiff bottlebrush structure (Fig. 7a, right, and Supplementary Video 2).

To become stiff, a bottlebrush has to be densely decorated by long side chains[45]. Short genes produce short nRNAs (Fig. 7b, left) corresponding to minute nRNPs, and hence cannot generate sufficient stiffness to prevent gene coiling even when a gene is highly transcribed (Fig. 7c). In contrast, long genes produce long nRNAs corresponding to large nRNPs inducing stiffness and gene expansion. Large nRNPs can be formed on genes with either multiple long exons (Fig. 7b, right), or long introns (Fig. 7b, middle). The first case is exemplified by the *Ttn* gene generating a 102-kb-long mRNA: the *cis*-to-*trans* ratio curve for *Ttn* displays an asymmetrical drop towards the 3′ end (Fig. 6f), indicating a stronger exclusion of the 3′ gene end from the harbouring locus. The second case is exemplified by the *Tg* gene with a 54-kb-long intron, over which mRNAs grow from ca. 6 kb to 60 kb, forming a signal gradient (Fig. 7d, top left). The intron displays a disproportionately larger extension compared with the rest of the gene (Fig. 7d, bottom and top right), consistent with greater stiffness. The strand opposite to *Tg* intron 41 includes the *Sla* gene, which is not expressed in thyrocytes (Supplementary Fig. 5) and thus does not contribute to intron extension.

Next, we used splicing inhibition to increase the length of nRNAs experimentally, reasoning that this would lead to an increased size of nRNPs (Fig. 7e, top left), which in turn would increase stiffness and, consequently, TL expansion. We inhibited splicing with pladienolide B[46] in myoblasts and performed RNA FISH with probes for the intron-rich 5′ end and the exon-rich 3′ end of the *Cald1* gene (Fig. 7e, right). Despite the massive abortion of transcription during drug treatment, we indeed observed a 2.5-fold increase of the signal size corresponding to *Cald1* TL parts formed by the intron-rich region of the gene (Fig. 7e, bottom left).

Polymer modelling confirms stiffness of TLs. We next turned to polymer modelling, aiming to understand whether stiffening and lengthening of a highly transcribed gene can give rise to an extended TL. We simulated six territorial chromosomes confined to a spherical nucleus by initiating them from a mitotic-like conformation and letting them expand (Fig. 8a, see Methods and Extended Data Fig. 10). On each chromosome, we modelled a 100-kb-long region as the gene of interest (Fig. 8a,c, top) and explored at which simulation parameters we can recapitulate the biological observables for the *Tg* gene: the appearance of TLs, distances between flanks measured by microscopy and a dip in *cis*-to-*trans* contact ratio seen in Hi-C (Fig. 8b).

First, we found that increasing gene length (by 15-fold because of nucleosomal loss: a 1 kb monomer extends from 20 nm to 300 nm) can increase the inter-flank distance, but fails to produce visible TLs (genes appear as amorphous clouds) or the dip in *cis*-to-*trans* ratio (Fig. 8c, second from top). Next, we reasoned that, due to the dense decoration with nRNPs, a highly transcribed gene has an increased stiffness. When we increased the contour length by 3-fold (partial nucleosome loss) and altered the stiffness (bending energy) of the polymer, we found that a 12-fold increase in the stiffness reproduced the experimentally observed features of *Tg* TLs: the appearance of TLs, flank separation and a dip in the *cis*-to-*trans* ratio (Fig. 8c, second from bottom). Moreover, the model recapitulates the dynamics of TL flanks upon gene silencing (Extended Data Fig. 10b). To account for the gradual increase of the nRNA length along the gene, we introduced a tapered stiffness profile (from 1- to 24-fold), which caused visibly more curled 5′ end and stretched 3′ end (Fig. 8c, bottom), recapitulating the *Tg* TLs, and asymmetric *cis*-to-*trans* profile along the gene, in agreement with Hi-C for the *Ttn* gene (Fig. 6f).

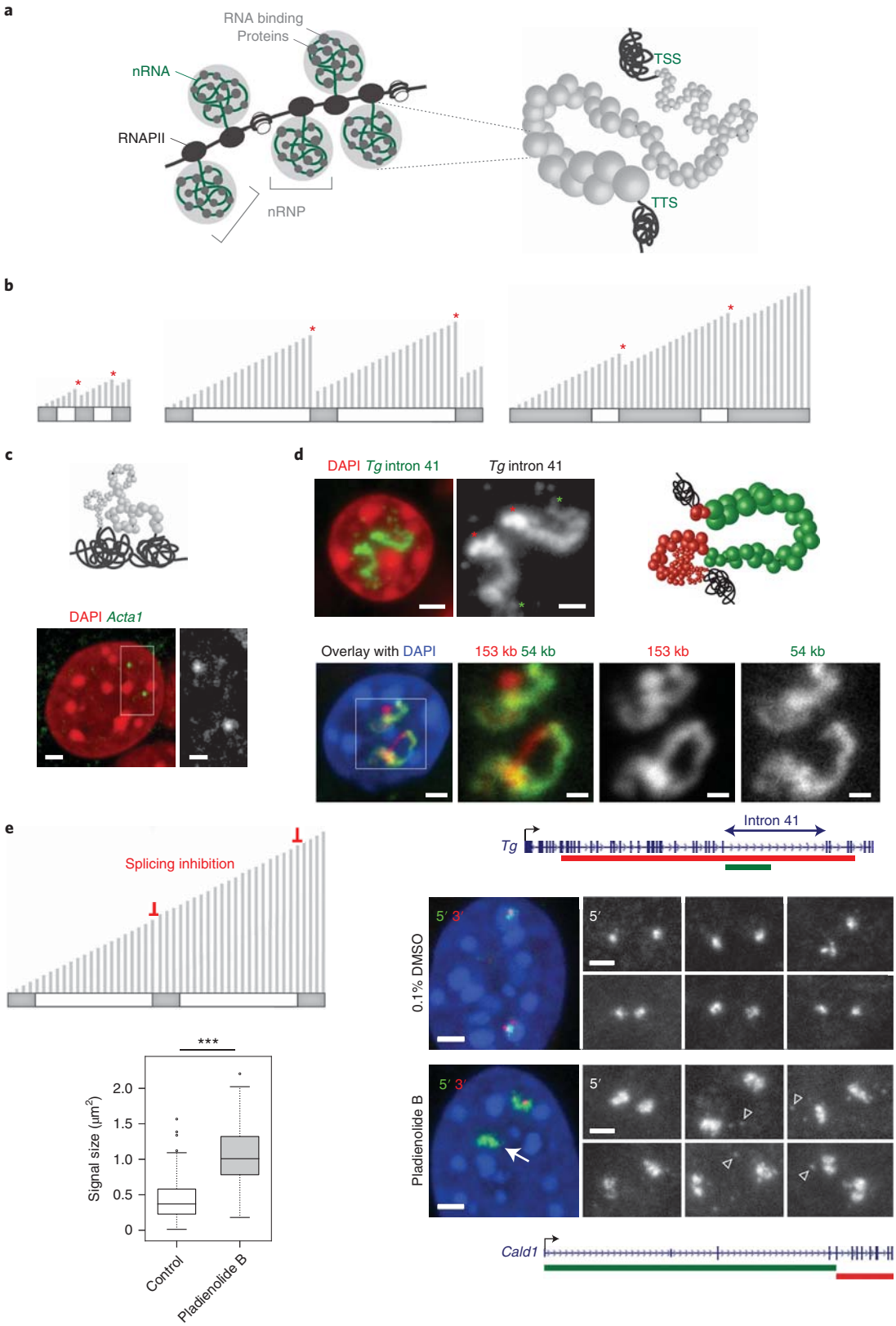
Consistent with experiments, the appearance of simulated TLs does not lead to insulation of long-range contacts and is not accompanied by major changes in Hi-C maps, except the *cis*-to-*trans* dip (Extended Data Figs. 9 and 10c). This suggests that chromosomes, simulated and real, are sufficiently flexible to accommodate the formation of TL without major reorganization.

Since the biological foundation of our stiffness hypothesis is the dense decoration of genes with nRNPs, we simulated a polymer grafted with side chains instead of imposing stiffness ad hoc. As expected from the theory[45], the region with side chains expands,

Fig. 7 | TLs of long highly expressed genes are stiff structures. a, Left: cartoon depicting the nRNPs formed during transcription elongation. Right: nRNPs formed on a long highly expressed gene are voluminous and densely decorate the gene axis leading to its stiffness and expansion. **b**, Schematics showing a short gene (left), a long gene with long introns (middle) and a long gene with long exons (right). Exons and introns are shown as dark-grey and white rectangles, respectively; transcripts are depicted as perpendicular light-grey lines of only half of the template length; splicing events are marked with red asterisks. **c**, Top: cartoon showing a short gene decorated by small RNPs allowing the gene axis to remain flexible and coiled. Bottom: example of highly expressed (4,360 TPM) but short (3 kb) *Acta1* gene exhibiting a small RNA FISH signal (bottom). **d**, The longest *Tg* gene intron 41 (54 kb) strongly expands and exhibits a gradient of nRNPs from 5′ (green asterisks) to 3′ (red asterisks) splice sites (top left). The expansion of the intron (green) is disproportional compared with the rest of the gene (green, bottom) as depicted in the cartoon (top right). Data represent two independent experiments. **e**, Top left: schematic showing the effect of splicing inhibition on the length of nRNAs. Right: comparison of *Cald1* TL size signal in control myoblasts (0.1% dimethylsulfoxide (DMSO)) and myoblasts treated with the splicing inhibitor pladienolide B (10 nM). RNA FISH signals of the 5′ and 3′ ends are shown in green and red, respectively. Grey-scale images show examples of the 5′ TL end. Pladienolide B causes massive abortion of transcription evident from the absence of the 3′ signal in a large proportion of *Cald1* alleles (arrow) and accumulation of nucleoplasmic granules detected by the 5′ probe (arrowheads). Bottom left: despite transcription abortion, the size of RNA FISH signals is increased 2.5-fold, indicating larger expansion of the TLs ($n = 118$ and 123 alleles in DMSO and pladienolide B treated cells, respectively, collected from 150 myoblasts across two independent experiments), $^{***}P = 2.788 \times 10^{-25}$, two-sided Wilcoxon rank sum test). Boxplots show the median (horizontal line), 25th to 75th percentiles (boxes) and 90th percentile (whiskers) of the group. Scale bars: 2 µm; in close-ups, 1 µm.

its flanks separate and the backbone becomes progressively stiffer with gradual increase of the side chain length (Fig. 8d). Computed *cis*-to-*trans* profiles, however, did not have a clear dip indicating the lack of externalization, probably because other factors underlying chromosomal territoriality were not taken into account by modelling.

Importantly, simulations show that a sufficiently dense polymer system (nuclear chromatin) is not confining and allows large-scale relocalization of a long polymer loop (transcribed gene) subject to a sufficiently low force due to stiffening. In conclusion, our simulations serve as a proof of principle that increased stiffness can explain TL formation.



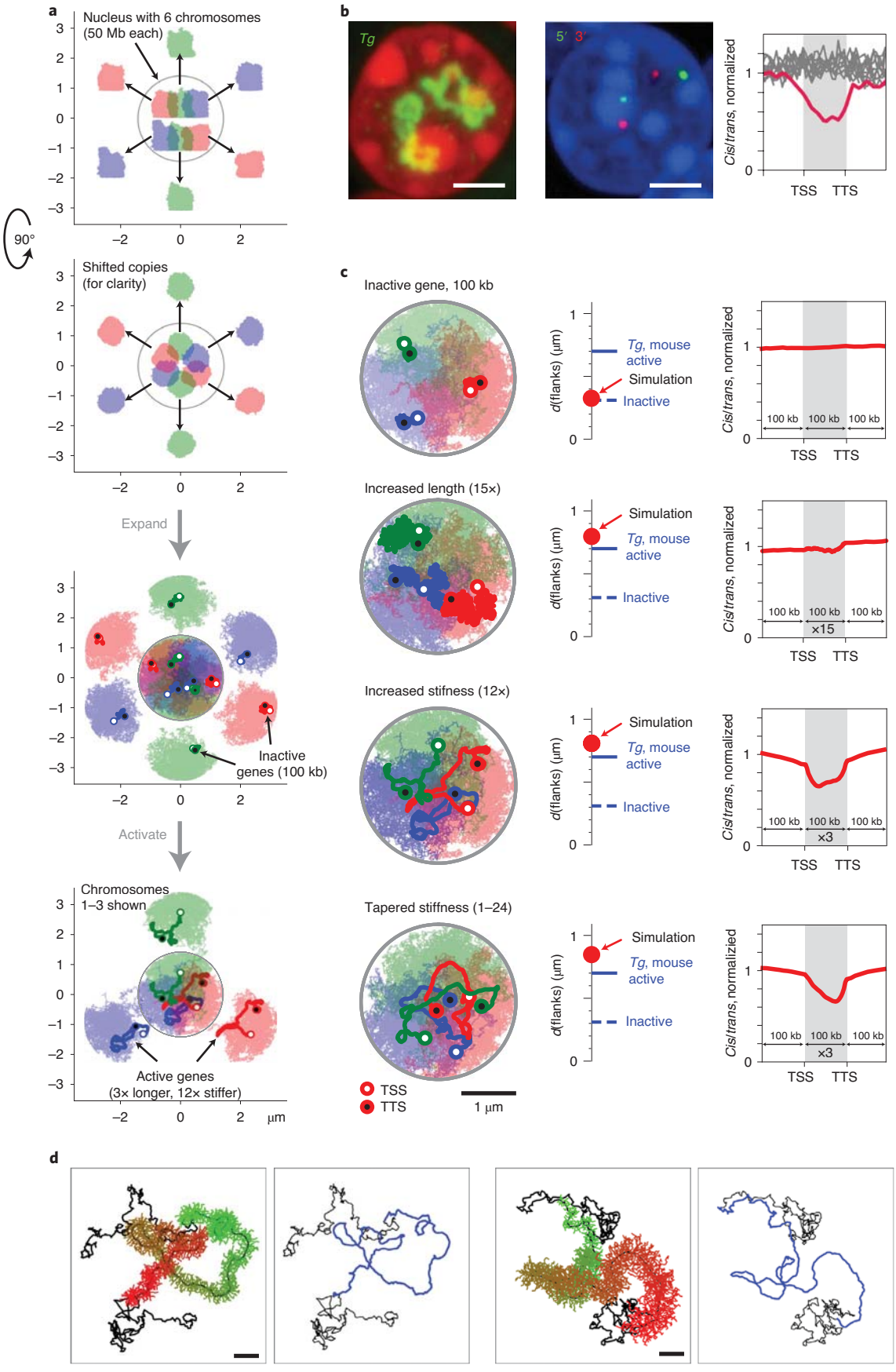

a
Nucleus with 6 chromosomes
(50 Mb each)
90°
Shifted copies
(for clarity)
Expand
Inactive
genes (100 kb)
Activate
Chromosomes
1–3 shown
Active genes
(3× longer, 12× stiffer)
μm
b
Tg
5′ 3′
Cis/trans, normalized
TSS TTS
c
Inactive gene, 100 kb
Increased length (15×)
Increased stifness (12×)
Tapered stiffness (1–24)
d(flanks) (μm)
Tg, mouse
active
Simulation
Inactive
100 kb 100 kb 100 kb
×15
×3
TSS
TTS
1 μm
d

Fig. 8 | Polymer modelling of transcribed genes. a, Polymer simulation setup. 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. Then polymers were expanded and, for each polymer, a 100 kb gene of interest was assigned (depicted with a darker colour). In the rest of the simulations, only three chromosomes are shown (bottom). **b**, Biological variables used to verify TL models: appearance of *Tg* TLs, inter-flank distances measured by microscopy and *cis*-to-*trans* contact ratios calculated from Hi-C maps; scale bar, 2 µm. **c**, Top: genes of 100 kb are compact structures and do not exhibit a dip in *cis*-to-*trans* contact ratio compared with the surrounding fibre in the inactive state. Second from top: a mere increase in gene contour length (15-fold) leads to a bigger TL volume but does not exhibit a clearly discernible contour or a dip in *cis*-to-*trans* contact ratio. Second from bottom: increased stiffness (12-fold in the bending energy) together with a moderate increase in contour length (3-fold) shows that TLs expand substantially and exhibit a clear contour and a dip in the *cis*-to-*trans* contact ratio. Bottom: a gradual increase in stiffness along the gene (on average 12-fold) leads to larger inter-flank distances (d(flanks)), more coiling of the 5′ compared with 3′ gene ends and an asymmetric dip in *cis*-to-*trans* contact ratio with a steeper slope at the 3′ end. Scale bar, 1 µm. **d**, Side chains as the source of increased stiffness. Left: a gene with 300 monomers, each decorated with three side chains made of five monomers (coloured from green to red), is shown together with the 300 monomer flanks (black). Right: side chain length increases gradually from 1 to 15 monomers. Right panels show the backbone of the simulated genes (blue) from the left panels. Scale bars, 0.2 µm.

Discussion

Our study focuses on the structure and spatial arrangement of individual transcribed genes in differentiated postmitotic cells and demonstrates that the sole process of transcription can have a profound effect on the spatial organization of gene loci by forming extended TLs, although functional implications of TLs remain to be understood. To enable light microscopy observations, we used long highly expressed genes as a model and demonstrated that the selected genes form open-ended TLs visually resembling lateral loops of lampbrush[7,22] and puffs of polytene chromosomes[8]. Importantly, TLs are formed as a result of transcription and as such should not be confused with enhancer–promoter interactions or loops extruded by SMC (structural maintenance of chromosomes) complexes[47].

The sequential patterns of exon and intron labelling along TLs strongly indicate that RNAPIIs move along the gene axis and carry nRNAs undergoing splicing (Supplementary Video 2). This mechanism is drastically different from the model of transcription factories with immobilized RNAPIIs and nRNAs extruded in a spot[12,13]. TLs, however, do not contradict the notion of transcription condensates[2,48–50] that form and dissolve at different transcription stages[51,52], although present at much smaller scales (submicrometres versus micrometres).

We show that transcription dynamically modifies the chromosomal locus harbouring a TL: transcription activation causes divergence of flanks and gene externalization from chromosome territory, whereas transcription silencing causes gene body condensation and convergence of its flanks. Separation of TL flanks argues against the proclaimed necessity of TSS–TTS association for maintenance of a high transcription level[53–55]. Moreover, the complicated geometry of highly extended TLs is not compatible with the idea of a perpetual contact between a promoter or an enhancer–promoter complex with the gene body during transcription[56,57].

The spreading of TLs over euchromatic nuclear areas by extending away from harbouring chromosomes and even breaking them apart challenges the significance of chromosome territoriality in transcription regulation[58]. A greater number of Hi-C *trans* contacts for both euchromatin in general and long highly expressed genes in particular indicates that expressed genes can extensively interact with the surrounding euchromatin. Taken together, these data argue that chromosome territoriality is not an essential functional feature of the interphase nucleus, as was recently demonstrated[52], but probably a mere consequence of the last mitosis[5]. Furthermore, the observed repositioning of a gene by several micrometres, together with polymer simulations, suggest that the nuclear environment is not confining but allows large-scale chromosomal movements, in agreement with recent chromosome micromanipulation[59] and arguing against proposals that interphase chromatin is a gel[60] or a solid[61].

We present evidence that highly expressed genes can generate a mechanical force sufficient to expand and relocalize them in nuclear space. We rule out the loss of histones as a possible mechanism as it causes gene lengthening but not externalization and TL formation. We suggest that TLs are characterized by increased stiffness caused by a dense decoration of the gene with progressively growing nRNPs. The hypothesis is supported by differential extension of TL regions with different sizes of nRNAs and by modelling a transcribed gene as a region of increased stiffness. This hypothesis also explains why highly expressed but short genes do not form resolvable TLs: a short gene lacks long introns or exons and is thus decorated by small nRNPs permitting gene-axis coiling (Fig. 7).

In conclusion, we argue that, while formation of microscopically resolvable TLs can be specific for long and highly expressed genes (Fig. 4), the mechanisms underlying the formation of such loops can be a general aspect of eukaryotic transcription (Supplementary Video 2).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests, and statements of data and code availability are available at https://doi.org/10.1038/s41556-022-00847-6.

Received: 9 April 2021; Accepted: 10 January 2022;
Published online: 17 February 2022

References

1. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* **21**, 71–87 (2020).
2. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
3. Herzel, L., Straube, K. & Neugebauer, K. M. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.* **28**, 1008–1019 (2018).
4. Feodorova, Y., Falk, M., Mirny, L. A. & Solovei, I. Viewing nuclear architecture through the eyes of nocturnal mammals. *Trends Cell Biol.* **30**, 276–289 (2020).
5. Solovei, I., Thanisch, K. & Feodorova, Y. How to rule the nucleus: divide et impera. *Curr. Opin. Cell Biol.* **40**, 47–59 (2016).
6. van Steensel, B. & Belmont, A. S. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell* **169**, 780–791 (2017).
7. Macgregor, H. C. *An Introduction to Animal Cytogenetics* (Chapman & Hall, 1993).
8. Bjork, P. & Wieslander, L. The Balbiani ring story: synthesis, assembly, processing, and transport of specific messenger RNA–protein complexes. *Annu. Rev. Biochem.* **84**, 65–92 (2015).
9. Chakalova, L. & Fraser, P. Organization of transcription. *Cold Spring Harb. Perspect. Biol.* **2**, a000729 (2010).
10. Osborne, C. S. et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36**, 1065–1071 (2004).
11. Schoenfelder, S. et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* **42**, 53–61 (2010).

12. Cook, P. R. The organization of replication and transcription. *Science* **284**, 1790–1795 (1999).
13. Papantonis, A. & Cook, P. R. Transcription factories: genome organization and gene regulation. *Chem. Rev.* **113**, 8683–8705 (2013).
14. Mateo, L. J. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
15. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
16. Rodermund, L. et al. Time-resolved structured illumination microscopy reveals key principles of Xist RNA spreading. *Science* **372**, (2021).
17. Morgan, G. T. Imaging the dynamics of transcription loops in living chromosomes. *Chromosoma* **127**, 361–374 (2018).
18. Schermelleh, L., Heintzmann, R. & Leonhardt, H. A guide to super-resolution fluorescence microscopy. *J. Cell Biol.* **190**, 165–175 (2010).
19. Jjingo, D., Huda, A., Gundapuneni, M., Marino-Ramirez, L. & Jordan, I. K. Effect of the transposable element environment of human genes on gene length and expression. *Genome Biol. Evol.* **3**, 259–271 (2011).
20. Hutchison, N. Lampbrush chromosomes of the chicken, *Gallus domesticus*. *J. Cell Biol.* **105**, 1493–1500 (1987).
21. Miller, O. L. & Beatty, B. R. Visualization of nuclear genes. *Science* **164**, 955–957 (1969).
22. Kaufmann, R., Cremer, C. & Gall, J. G. Superresolution imaging of transcription units on newt lampbrush chromosomes. *Chromosome Res.* **20**, 1009–1015 (2012).
23. Carmo-Fonseca, M. & Kirchhausen, T. The timing of pre-mRNA splicing visualized in real-time. *Nucleus* **5**, 11–14 (2014).
24. Ansari, A. & Hampsey, M. A role for the CPF 3′-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* **19**, 2969–2978 (2005).
25. Larsson, A. J. M. et al. Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).
26. Tantale, K. et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7**, 12248 (2016).
27. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
28. Bensaude, O. Inhibiting eukaryotic transcription: which compound to choose? How to evaluate its activity? *Transcription* **2**, 103–108 (2011).
29. Mahy, N. L., Perry, P. E., Gilchrist, S., Baldock, R. A. & Bickmore, W. A. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J. Cell Biol.* **157**, 579–589 (2002).
30. Brown, J. M. et al. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083–1097 (2008).
31. Chambeyron, S., Da Silva, N. R., Lawson, K. A. & Bickmore, W. A. Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development* **132**, 2215–2223 (2005).
32. Mahy, N. L., Perry, P. E. & Bickmore, W. A. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol.* **159**, 753–763 (2002).
33. Volpi, E. V. et al. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* **113**, 1565–1576 (2000).
34. Williams, R. R., Broad, S., Sheer, D. & Ragoussis, J. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp. Cell. Res.* **272**, 163–175 (2002).
35. Abramo, K. et al. A chromosome folding intermediate at the condensin-to-cohesin transition during telophase. *Nat. Cell Biol.* **21**, 1393–1402 (2019).
36. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98 (2012).
37. Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376 e316 (2018).
38. Hsieh, T. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553 (2020).
39. Banigan, E. J. & Mirny, L. A. The interplay between asymmetric and symmetric DNA loop extrusion. eLife 9, (2020).
40. Brandao, H. B. et al. RNA polymerases as moving barriers to condensin loop extrusion. *Proc. Natl Acad. Sci. USA* **116**, 20489–20499 (2019).
41. Banigan, E. J. et al. Transcription shapes 3D chromatin organization by interacting with loop-extruding cohesin complexes. https://doi.org/10.1101/2022.01.07.475367 (2022).
42. Muller-McNicoll, M. & Neugebauer, K. M. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat. Rev. Genet.* **14**, 275–287 (2013).
43. Olins, A. L. & Olins, D. E. Spheroid chromatin units (v bodies). *Science* **183**, 330–332 (1974).
44. Liu, X., Farnung, L., Wigge, C. & Cramer, P. Cryo-EM structure of a mammalian RNA polymerase II elongation complex inhibited by alpha-amanitin. *J. Biol. Chem.* **293**, 7189–7194 (2018).
45. Paturej, J., Sheiko, S. S., Panyukov, S. & Rubinstein, M. Molecular structure of bottlebrush polymers in melts. *Sci. Adv.* **2**, e1601478 (2016).
46. Kotake, Y. et al. Splicing factor SF3b as a target of the antitumor natural product pladienolide. *Nat. Chem. Biol.* **3**, 570–575 (2007).
47. Mirny, L. A. & Solovei, I. Keeping chromatin in the loop(s). *Nat. Rev. Mol. Cell Biol.* **22**, 439–440 (2021).
48. Cho, W. K. et al. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).
49. Guo, Y. E. et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548 (2019).
50. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
51. Cisse, I. I. et al. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664–667 (2013).
52. Henninger, J. E. et al. RNA-mediated feedback control of transcriptional condensates. *Cell* **184**, 207–225 e224 (2021).
53. Hampsey, M., Singh, B. N., Ansari, A., Laine, J. P. & Krishnamurthy, S. Control of eukaryotic gene expression: gene loops and transcriptional memory. *Adv. Enzyme Regul.* **51**, 118–125 (2011).
54. Martin, M., Cho, J., Cesare, A. J., Griffith, J. D. & Attardi, G. Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis. *Cell* **123**, 1227–1240 (2005).
55. Singh, B. N. & Hampsey, M. A transcription-independent role for TFIIB in gene looping. *Mol. Cell* **27**, 806–816 (2007).
56. Lee, K., Hsiung, C. C., Huang, P., Raj, A. & Blobel, G. A. Dynamic enhancer–gene body contacts during transcription elongation. *Genes Dev.* **29**, 1992–1997 (2015).
57. Zheng, M. et al. Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).
58. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
59. Keizer, V. I. P. et al. Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.20.439763 (2021).
60. Khanna, N., Zhang, Y., Lucas, J. S., Dudko, O. K. & Murre, C. Chromosome dynamics near the sol-gel phase transition dictate the timing of remote genomic interactions. *Nat. Commun.* **10**, 2771 (2019).
61. Strickfaden, H. et al. Condensed chromatin behaves like a solid on the mesoscale in vitro and in living cells. *Cell* **183**, 1772–1784 e1713 (2020).

Methods

GTEx data analysis. RNA-seq data on human tissues were downloaded from the website of the GTEx portal (https://gtexportal.org/home/datasets, version 7, median gene level TPMs). The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, by the National Cancer Institute, the National Human Genome Research Institute, the National Heart, Lung, and Blood Institute, the National Institute on Drug Abuse, the National Institute of Mental Health and the National Institute of Neurological Disorders and Stroke. GENCODE gene annotation version v19 (human) was downloaded from https://www.gencodegenes.org/. Using R (version 3.5.3, https://www.R-project.org/), mitochondrial genes were excluded from the analysis and only protein-coding genes with a known gene status were selected. 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 ID (Supplementary Table 1).

RNA-seq. Total RNA from cells or tissues was isolated using the NucleoSpin RNA Kit (Macherey-Nagel) according to the manufacturer's instructions. Digital gene expression libraries for RNA-seq were produced using a modified version of single-cell RNA barcoding sequencing optimized to accommodate bulk cells[62,63]. After digestion, purification and reverse transcription, the segments were subjected to Illumina sequencing adaptors. During the library preparation PCR, 3′ ends were enriched with a custom P5 primer (P5NEXTPT5, Integrated DNA 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 complementary DNA fragment. As multiple libraries were sequenced in parallel on the flow cell, an additional eight-base i7 barcode read was performed.

RNA-seq libraries were processed and mapped to the mouse genome (mm10) using the zUMIs pipeline[64]. Unique molecular identifier count tables were filtered for low counts using HTSFilter[65]. GENCODE gene annotation version vM11 (mouse) was downloaded from https://www.gencodegenes.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 (Supplementary Table 2).

Cell culture. The mouse myoblast cell line Pmi28 was grown in Nutrient Mixture F-10 Ham supplemented with 20% foetal bovine serum and 1% penicillin/streptomycin at 37 °C and 5% CO_2. For differentiation, cells were seeded at a density of 8×10^4 cells per cm² 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 d. A sufficient number of myotubes typically formed after 3–4 d.

Manipulation of transcription and splicing. Transcription stimulation was performed using a system previously described[27]. Pmi28 mouse myoblasts were transfected with PB-TRE-dCas9-VPR (Addgene, ID 63800) and PiggyBac transposase (System Biosciences, #PB200PA-1) expression plasmids at a ratio of 3:1 with Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions. Clones that stably integrated dCas9-VPR were selected with a medium containing 50 μg ml^{-1} hygromycin B (Sigma-Aldrich). Single cells were sorted by fluorescence-activated cell sorting (BD Biosciences Aria II Sorter) and cultured for one week in the presence of hygromycin B. dCas9-VPR expression was induced with 1 μg ml^{-1} doxycyclin for 50 h. Individual clones were tested for dCas9 expression by western blot using a Cas9 antibody (Clontech, dilution 1:1,000). For guide RNA (gRNA) transfection, 3×10^5 cells were seeded in six wells on coverslips and transfected with 2.5 μg of a mix of six different gRNA plasmids per target gene (Supplementary Table 3) at equimolar amounts with Lipofectamine 3000 according to the manufacturer's instructions. To monitor 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 ref. [66]. Then U6-gRNA-GFP-H2A was added to the gRNA cocktail. dCas9-VPR expression was induced after 12 h by addition of 1 μg ml^{-1} doxycyclin. After 50 h of doxycycline treatment, the expression of full-length *Ttn* mRNA was verified by reverse-transcription PCR on sequences located at the 5′ and 3′ ends of the full-length *Ttn* transcript with the primers indicated in Supplementary Table 3. Afterwards, the relative expression level of *Ttn* was assessed by quantitative real-time PCR on a LightCycler 480 II (Roche) using LightCycler 480 SYBR Green (Roche) according to the manufacturer's instructions with the following primers: Ttn fwd (5′-GCCGCGCTAGATTGATGATC-3′) and Ttn rev (5′-TCTCGGCTGTCACAAGAAGCT-3′); see also Supplementary Table 4. All generated plasmids are available upon request.

For transcription inhibition in myotubes, differentiation medium was supplemented with 10 μg ml^{-1} actinomycin D (Sigma-Aldrich) for 12 or 24 h, or 25 μg ml^{-1} α-amanitin (Sigma-Aldrich) for 24 h, or 100 μM DRB (Sigma-Aldrich) for 3 or 6 h. Fixation timepoints of cells under DRB treatment or after DRB removal included 2 min, 25 min, 50 min, 75 min and 3 h.

For splicing inhibition in myoblasts, the culture medium was supplemented with 10 nM of pladienolide B (Cayman Chemical) for 4 h before fixation.

Preparation of samples for ChIP–seq and Hi-C. Pmi28 myoblasts were detached using 0.25% trypsin/1 mM EDTA for 3 min and pelleted. For native chromatin immunoprecipitation followed by sequencing (ChIP–seq), the pellet (roughly corresponding to 15×10^6 myoblasts) was directly snap-frozen in liquid nitrogen and kept at −80 °C. For Hi-C, cells were resuspended in medium to a final volume of 10 ml, and 666 μl of 16% formaldehyde was added (resulting in a final concentration of 1% formaldehyde); the tube was briefly shaken and then rotated for exactly 10 min at 21 °C. To quench the reaction, 593 μl of 2.5 M glycine was 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 room temperature and the supernatant was discarded. The pellets were snap-frozen in liquid nitrogen and stored at −80 °C.

Myotube samples were collected as follows 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 to allow the remaining myoblasts to re-adhere to the dish bottom. Since myotubes do not re-adhere, they were collected with the supernatant. For native ChIP–seq, the pellet (corresponding to 10×10^6 myotubes) was snap-frozen in liquid nitrogen and kept at −80 °C. For Hi-C, myotubes were fixed and frozen in the same way as myoblasts. We estimated that our myotube samples contained 10–20% myoblasts.

Mouse thyroids for Hi-C were prepared as follows: thyroid glands from three mice were dissected within 5–10 min and stored in medium containing thyroid stimulation hormone (TSH) (DMEM/F-12; 20% FCS; penicillin/streptomycin; 2 mM L-glutamine; 40 μg ml^{-1} vitamin C; 1 U ml^{-1} TSH) in a cell culture incubator at 37 °C. Glands were then processed one by one. The two lobes of each gland were cleaned from attached neighbouring tissues 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 (five to six from each lobe) and transferred to a 2 ml tube using low-retention tips. The tube was centrifuged at 800*g* for 5 min at room temperature, and the supernatant was discarded. Tissue was resuspended in 1.5 ml of DMEM without FCS. Then, 100 μl of 16% formaldehyde was added (resulting in a final concentration of 1% formaldehyde); the tube was briefly shaken and then rotated for exactly 10 min at 21 °C. To quench the reaction, 89 μl of 2.5 M glycine was added; the tube was briefly shaken and left rotating for 5 min. The tube was moved to ice and kept there for 0.5 h. After quenching, tubes were centrifuged for 10 min at 2,000*g* and the supernatant was discarded. The pellets were snap-frozen in liquid nitrogen and kept at −80 °C. The same procedure was repeated with three other mice so that each biological replicate consisted of two tubes each containing cells from six thyroid lobes. We estimated that mouse thyroid gland contains 400,000–1,000,000 cells (thus no less than 3,000,000 cells per biological replicate) and that thyrocytes constitute ca. 60% of thyroid cells.

For smooth muscles, one mouse bladder per sample was used. The tissue was minced and fixed as described for thyroid. We estimated that bladder samples contained 40–50% of smooth muscle cells.

Native ChIP–seq. Native ChIP–seq was performed as previously described[67,68] with the following alterations: immunoprecipitations (IPs) were carried out with S1 mononucleosomes derived from Pmi28 myoblasts or myotubes. Experiments were performed in independent duplicates. Per one IP reaction (500 μl), mononucleosomes from 5×10^6 cells were incubated with 5 μg of antibody (anti-RNAPII CTD repeat YSPTSPS, 8WG16, Abcam ab817, PRID AB_306327) in rotating tubes at 4 °C overnight. Then the antibody was post-coupled to magnetic beads (Dynabeads Protein G, Invitrogen) at 4 °C for 3 h and the eluted IP DNA was phenol/chloroform/isoamylalcohol extracted using MaXtract high-density tubes (Qiagen) and ethanol-precipitated. 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. ChIP–seq libraries were sequenced on an Illumina HiSeq 1500 platform using 50 bp single-end nucleotide sequencing.

Hi-C. Three Hi-C biological replicates were prepared for each of the studied cell types. Sample sizes were estimated as follows: 15×10^6 of myoblasts, 10×10^6 of myotube nuclei and 3×10^6 of thyroid cells (thyroids of six mice); the number of cells in a bladder was not estimated. Hi-C was performed as described by ref. [69]. Myoblasts and myotubes were split into three or two subsamples, respectively, each containing 5×10^6 cells, which were treated as separate samples until the DNA purification step when they were merged into one biological replicate. Chromatin was digested with the restriction enzyme *DpnII* (New England Biolabs) at 37 °C overnight. Chromatin sonication and size selection were performed with some

modifications to allow retrieval of a higher DNA quantity. Hi-C libraries were sequenced using 50 bp paired-end sequencing on an Illumina HiSeq 4000. For the number of reads obtained in each Hi-C experiment, see Supplementary Table 4.

Hi-C data processing. Fastq files were mapped to the mm10 genome using the distiller pipeline (http://github.com/mirnylab/distiller-nf). Hi-C interaction data were stored in the cooler format[70] Cooler: scalable storage for Hi-C data and other genomically labelled arrays (https://github.com/mirnylab/cooler). We always applied iterative correction to our data to remove biases as much as possible[71]. For the reproducibility of the replicates, see Supplementary Fig. 4. *Cis*-to-*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*-to-*trans* contact frequency ratios.

For a comparison plot of the *cis*-to-*trans* ratios across a number of genes the *x* axis for each gene was rescaled so that TSS and TTS align. Interpolation was used to obtain the same number of points in each gene. The *cis*-to-*trans* ratios were normalized to regions outside the genes (to be precise, the region from TSS 3′ gene length to TSS 0.5′ gene length and TTS + 0.5′ gene length to TTS + 3′ gene length is used to normalize *cis*-to-*trans* profile to unity). This is done to highlight local changes in the *cis*-to-*trans* ratio associated with genes while suppressing longer-range variations. For statistical validation, *cis*-to-*trans* ratios were compared for every 32 kb bin of target and control genes (see Supplementary Table 2 for the list of control genes). The statistical test was a pairwise Mann–Whitney rank test for smaller values in the target group with significance $P < 0.01$.

Mice and tissue sampling. CD-1 mice were purchased from Charles River Laboratories, housed in individual cages with free access to food and water on a 12 h–12 h light–dark cycle 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. Mice were sacrificed by cervical dislocation after IsoFlo (Isofluran, Abbott) narcosis.

Isolated tissues were washed with PBS and then fixed with 4% paraformaldehyde (Carl Roth) solution in PBS for 12–20 h. After fixation, mouse tissues were washed with PBS, cryoprotected in a series of sucrose and embedded in Tissue-Tek OCT compound freezing medium (Sakura). Blocks were stored at −80 °C before cutting into 16–20 µm sections using a cryostat (Leica CM3050S). Cryosections were collected on Superfrost Plus slides (Thermo Scientific) and stored at −80 °C before use.

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

FISH probes. *BAC-derived FISH probes.* BAC clones encompassing the desired genomic region (Supplementary Table 3) were selected using the UCSC genome browser (https://genome.ucsc.edu/) and purchased from BACPAC Resources (Children's Hospital Oakland) as agar stabs (https://bacpacresources.org/). BACs were purified via standard alkaline lysis or the NucleoBond Xtra Midi Kit (Macherey-Nagel), followed by amplification with the GenomiPhi Kit (GE Healthcare) according to the manufacturer's instructions. BAC probes were labelled either directly with fluorophores (using fluorophore dUTPs), or with haptens (biotin dUTP and digoxigenin dUTP) by nick translation. Labelling of dUTPs was performed as described in ref. [72].

Chromosome paints. Whole chromosome paints were a gift from N. Carter and J. Wienberg (Cambridge). Paints were amplified and labelled with biotin or digoxigenin via degenerate oligonucleotide-primed PCR with a partially degenerate oligonucleotide primer (6 MW primer: 5′-CCG ACT CGA GNN NNN NAT GTG G-3′, Eurogentec)[72].

Tg exon probes. Probes that specifically label exons 2–12 and 33–47 of *Tg* were generated by standard PCR using a cDNA clone (transOMIC technologies, BC111467) as a template. Primer sequences are listed in Supplementary Table 3. PCR-amplified DNA was directly labelled with fluorophores by nick translation.

Oligoprobes. Most of the oligoprobes were generated using the original SABER FISH protocol[73]. BED files for the genomic locations of the exons and introns of the target genes were acquired from the UCSC Genome Browser. The designed oligos were ordered as oligo pools from Integrated DNA Technologies with a scale of 50 pmol per oligo. All used probe oligos are listed in Supplementary Table 3. The oligos were remapped if necessary for multi-colour imaging and extended to ~500 nt using the previously described primer exchange reaction[74]. Finally, the probes were purified using PCR clean-up columns (Macherey-Nagel).

Oligoprobes for Tg intron 41. FISH probe generation was conducted as described previously for HD FISH[75]. The probe covering the first half of *Tg* intron 41 was amplified from genomic J1 ESC wt DNA. Primers to amplify fragments for FISH probes were designed with the HD FISH probe generator platform accessible via http://www.hdfish.eu/[75] and are listed in Supplementary Table 3. After amplification, DNA was labelled with Alexa594 or Alexa647 using Ulysis Nucleic Acid Labeling Kits (Invitrogen).

FISH. FISH on 3D-preserved cultured cells was performed as previously described[72,76,77]. For DNA FISH, cells were treated with 50 µg ml^{-1} RNase A at 37 °C for 1 h. 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. Hybridizations were typically carried out in a water bath at 37 °C for 2 d. ImmunoFISH was performed as previously described[76]. Briefly, after immunostaining, cells were post-fixed with 2% PFA/PBS for 10 min, incubated in 20% glycerol for 1 h and further treated according to the protocol for 3D FISH. RNA FISH with oligoprobes (SABER FISH) was performed by mounting SABER probe containing hybridization mixture on fixed cells and incubating O/N at 43 °C. The samples were washed with 2× SSC at 37 °C twice for 20 min and with 0.1× SSC at 60 °C for 5 min. Hybridized SABER probes were detected by incubating with 1 µM fluorescently labelled detection oligonucleotides in PBS for 1 h at 37 °C followed by washing with PBS for 10 min. FISH on cryosections was performed as previously described[78]. Simultaneous denaturation of probes and sections was carried out on a hot block at 80 °C for 3 min. For combined RNA FISH and immunostaining, the cryosections were first hybridized with probes, then permeabilized with 0.1% Triton X-100/PBS and incubated with both primary and secondary antibodies as described previously[79–81].

The primary antibodies used in immunoFISH experiments included rabbit anti-Cas9 (1:150, Clontech, cat. no. 632607), mouse anti-GFP (1:400, Roche, cat. no. 11814460001, PRID: AB_390913), mouse anti-RNAPII-Ser2ph (1:50, clone 26B5, gift from Dr. H. Kimura, Tokyo Institute of Technology) and rabbit anti-SON (1:50, Sigma, cat. no. HPA023535). Donkey anti-mouse Alexa555 (1:250, Invitrogen, cat. no. A31570) and donkey anti-rabbit Alexa488 (1:250, Invitrogen, cat. no. A21202) were used as secondary antibodies.

In all experiments, nuclei were counterstained with 2 µg ml^{-1} DAPI in PBS for 10 min (cells) or 30 min (sections) and Vectashield (Vector) was used as an antifade mounting medium. For stimulated emission depletion (STED) microscopy, nuclei were counterstained with SiR DNA and samples were mounted into Mowiol.

Microscopy. Confocal image stacks were acquired using a TCS SP5 confocal microscope (Leica) with 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. Axial chromatic shift correction, as well as building single grey-scale stacks, RGB stacks, RGB montages and RGB maximum intensity projections, was performed using ImageJ plugin StackGroom[82]. The plugin is available upon request. For deconvolution of confocal stacks, a Lightning module of SP8 confocal microscope was used. High-resolution microscopy of TLs after RNA FISH was performed using either OMX (DeltaVision OMX V3 3D-SIM microscope, Applied Precision Imaging, GE Healthcare) equipped with lasers for DAPI and red fluorescence or with 3D STED (Abberior Instruments) with excitation pulsed lasers 594 and 640 nm and pulsed depletion laser at 775 nm.

Image analysis. Inter-flank distances were measured using in-house semi-automated program. The signal spots in both channels were identified by detecting local minima in a Laplacian-of-Gaussian-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.

The number of alleles in which inter-flank distances were measured: for thyrocytes, $n = 203$ and for control epithelial cells, $n = 180$; for myotubes, $n = 117$ and for control myoblasts, $n = 63$; for myotubes after irreversible transcription inhibition with α-amanitin, $n = 78$ and actinomycin D, $n = 112$ and for control untreated myotubes, $n = 62$; for myotubes after reversible transcription inhibition with DRB for 6 h, $n = 113$, after DRB removal, $n = 87$ and for control untreated myotubes, $n = 116$.

Contour length of the *Tg* TL regions were measured on maximum-intensity projections of stacks after RNA FISH. Loop contours were manually traced using the *Segmented Line* tool in ImageJ. 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.

TL modelling. Chromosomes were modelled as polymers with each monomer representing 1 kb or roughly five nucleosomes. To fix the 3D size of a monomer,

literature values for the compaction ratio c of the chromatin fibre, that is, the number of base pairs per nanometre of contour length were used. Literature values vary greatly and are in the range $c = 25–150$ bp nm^{-1} (refs. [83–85]). Here we used $c = 50$ bp nm^{-1}. We used a monomer size of 20 nm, corresponding to five nucleosomes or 1 kb.

For the Kuhn length l_K (typical bending radius of the fibre), again, literature values vary greatly and are in the range of $l_K = 60–400$ nm[83–85]. 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 (Fig. 3). Those measurements suggest somewhat smaller Kuhn lengths between 25 nm and 80 nm. For our polymer model, we chose $l_K = 56$ nm or 2.8 kb.

Finally, we fixed the volume density of chromatin in the cell nucleus. It can be estimated from the genome size (~3 Gb for a single copy and ~6 Gb per nucleus) and the size of the nucleus. With a diameter of 8 µm, we get a volume density of 180 bp in a cube of size (20 nm)^3, or 0.18 monomers/(20 nm)^3, which equals 18%. A cryo-electron microscopy measurement[86] suggests volume densities around 30%. For our polymer simulations, we chose 0.2 beads per (20 nm)^3 or 20%.

To interpret simulations, the simulation time scale is converted to physical time. This conversion is done by comparing mean square displacements of simulation monomers with experimental data on tracked loci in yeast[87] and mammals[88] as described in ref. [89]. A conversion factor of 656 ps in simulations corresponding to 1 s real time was obtained.

Polymer simulations were performed using a Mirny lab written wrapper (available at https://github.com/mirnylab/openmm-polymer-legacy) around the open-source graphics processing unit-assisted molecular dynamics package Openmm (1,2). Polymers are represented as a chain of monomers with harmonic bonds, a repulsive excluded volume potential ($3\,k_BT$), and are simulated with a collision rate of 0.1 inverse picoseconds (see ref. [89] for details on simulation parameters and setup). We imposed an energetic cost of non-zero bond angles. Specifically, the potential energy for bond angles is $E(\alpha) = k\alpha^2 k_BT/2$. Unless otherwise mentioned (see below), we used $k = 1$ to obtain the aforementioned Kuhn length.

We included A/B compartmentalization and lamina attraction. A/B compartmentalization was obtained by adding a weak attraction between monomers of type B ($0.1\,k_BT$), which induces phase separation. The segmentation of our 50 Mb chromosomes into A and B compartmental segments is based on a random distribution. We used the same segmentation as in ref. [89] but scaled up by a factor of 2.5 to account for the reduced coarse graining. The mean compartmental segment size was 650 monomers or 650 kb. Furthermore, the association of B-type chromatin with the nuclear periphery was modelled by a weak attraction ($0.6\,k_BT$) of B-type monomers with the sphere boundary representing the nuclear envelope. We point out that our simulation results on TLs do not depend on the presence of A/B compartments and lamina attraction.

Six chromosomes of 50 Mb each (corresponding to 50,000 monomers each) were simulated and placed inside a sphere with a diameter of 2.84 µm (corresponding to 42 monomer diameters). Territorial chromosomes were generated by initiating them in a mitotic-like conformation and letting them 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. The above expansion from mitosis was repeated ten 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. Simulations of TLs were then run for 127 min of real time (5,000,000 simulation time units), unless otherwise mentioned. We performed 30 runs (thus using each initial conformation three times). In each run, 101 conformations were saved at equidistant timepoints.

To simulate the effect of a high expression of a long gene, a 100 kb region (100 monomers) on each simulated chromosome was assigned to the gene of interest. First, to study the effect of a mere increase in contour length, we increased the region assigned as the gene of interest to cover a 15-fold longer region, that is, 1,500 kb. The fibre properties—rigidity and monomer size—were not altered. Then, to study the effect of increased stiffness, in the 100 kb regions the bending rigidity of the fibre was increased. To simulate increased stiffness, the parameter k was increased from our default value of unity (see above) to the specified values, for example to 12. The 12-fold increase in the bending energy was chosen by sweeping a range of values and choosing one that best reproduces the experimentally measured end-to-end distance of TLs. The 'genes' were positioned in a region of A-type chromatin.

Simulations of genes with side chains were performed by attaching short side chains onto monomers of the gene backbones. The side chains were identical to the backbone in terms of monomer size, excluded volume potential and bond force. There was no bond angle energy within the side chains. We attached three side chains to each backbone monomer (other grafting densities yielded similar results). We used uniform side-chain length (five monomers per side chain, that is, 100 nm side-chain length) as well as side-chain length gradually increasing from 1 to 15 monomers along the gene backbone.

To obtain Hi-C maps from simulated data, polymer conformations were coarse grained by a factor of ten (that is, only every tenth monomer is considered) to reduce the size of the computed Hi-C matrix. A cut-off radius was then designed mimicking the crosslinking radius in an actual Hi-C experiment. The cut-off radius was ten monomer diameters. Hi-C maps were computed from conformations in the second half of each run. *Cis*-to-*trans* contact frequency ratio tracks were computed as for experimental Hi-C matrices. The *cis*-to-*trans* ratio tracks of all six simulated chromosomes were aggregated to improve statistics.

Statistics and reproducibility. The experiments shown in this study were performed as two to ten biologically independent experiments, and no inconsistent results were observed. Data plotted as boxplots indicate the 25th and 75th percentiles, with the whiskers showing the minima and maxima (10th and 90th percentiles), black circles indicating the outliers and the horizontal line showing the median, if not stated otherwise. Some data are plotted in bar graphs as the mean ± s.d. If not stated otherwise, statistical testing was performed using the two-sided Wilcoxon rank sum test. Details of the particular statistical analyses used, precise P values, statistical significance, number of independent biological replicates and sample sizes for all of the graphs are indicated in the figures or figure legends. No data were excluded. For the number of acquired nuclei after FISH experiments, see Supplementary Table 4.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Hi-C data have been uploaded to Gene Expression Omnibus (GEO) and are available under accession GSE150704. ChIP–seq and RNA-seq data are available at ArrayExpress (EMBL-EBI) under accession E-MTAB-9060 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9060/). Previously published reference genome mm10 and gene annotation of the C57BL/6 J strain were downloaded from the Ensemble database (version GRCm38, release 74). Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding authors on reasonable request.

Code availability

The used code for the measurement of flank distances is available at https://github.com/hoerldavid/fish_analysis; code used for the analysis of Hi-C data (Cooler, Cooltools, Distiller, Pairtools) is available at https://github.com/open2c; the code used for polymer simulations is available at https://github.com/mirnylab/openmm-polymer-legacy.

References

62. Bagnoli, J. W. et al. Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nat. Commun.* **9**, 2937 (2018).
63. Ziegenhain, C. et al. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643 e634 (2017).
64. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs—a fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* **7**, (2018).
65. Rau, A., Gallopin, M., Celeux, G. & Jaffrezic, F. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* **29**, 2146–2152 (2013).
66. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
67. Link, S. et al. PWWP2A binds distinct chromatin moieties and interacts with an MTA1-specific core NuRD complex. *Nat. Commun.* **9**, 4300 (2018).
68. Punzeler, S. et al. Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation. *EMBO J.* **36**, 2263–2279 (2017).
69. Belaghzal, H., Dekker, J. & Gibcus, J. H. Hi-C 2.0: an optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* **123**, 56–65 (2017).
70. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
71. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
72. Cremer, M. et al. Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol. Biol.* **463**, 205–239 (2008).
73. Kishi, J. Y. et al. SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nat. Methods* **16**, 533–544 (2019).
74. Kishi, J. Y., Schaus, T. E., Gopalkrishnan, N., Xuan, F. & Yin, P. Programmable autonomous synthesis of single-stranded DNA. *Nat. Chem.* **10**, 155–164 (2018).
75. Bienko, M. et al. A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. *Nat. Methods* **10**, 122–124 (2013).
76. Solovei, I. & Cremer, M. 3D-FISH on cultured cells combined with immunostaining. *Methods Mol. Biol.* **659**, 117–126 (2010).
77. Solovei, I. et al. in *FISH: A Practical Approach* (eds Beatty, B. et al.) 119–157 (Oxford Univ. Press, 2002).
78. Solovei, I. Fluorescence in situ hybridization (FISH) on tissue cryosections. *Methods Mol. Biol.* **659**, 71–82 (2010).

79. Eberhart, A. et al. Epigenetics of eu- and heterochromatin in inverted and conventional nuclei from mouse retina. *Chromosome Res.* **21**, 535–554 (2013).
80. Eberhart, A., Kimura, H., Leonhardt, H., Joffe, B. & Solovei, I. Reliable detection of epigenetic histone marks and nuclear proteins in tissue cryosections. *Chromosome Res.* **20**, 849–858 (2012).
81. Solovei, I. et al. LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152**, 584–598 (2013).
82. Walter, J. et al. Towards many colors in FISH on 3D-preserved interphase nuclei. *Cytogenet Genome Res.* **114**, 367–378 (2006).
83. Bystricky, K., Heun, P., Gehlen, L., Langowski, J. & Gasser, S. M. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl Acad. Sci. USA* **101**, 16495–16500 (2004).
84. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
85. Rippe, K. Making contacts on a nucleic acid polymer. *Trends Biochem. Sci* **26**, 733–740 (2001).
86. Ou, H. D. et al. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, (2017).
87. Hajjoul, H. et al. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res.* **23**, 1829–1838 (2013).
88. Lucas, J. S., Zhang, Y., Dudko, O. K. & Murre, C. 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell* **158**, 339–352 (2014).
89. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl Acad. Sci. USA* **115**, E6697–E6706 (2018).

Acknowledgements

We are grateful to D. Hörl and J. Ryan for the help with ImageJ plugins and programming. We thank A. Maiser and K. Brandstetter for the help with high-resolution microscopy. We acknowledge J. Bates, D. Eick, P. Becker, M. Carmo-Fonseca, A. Olins and D. Olins and H.C.Macgregor for fruitful and insightful discussions. We thank D. Kralev and T. Suzuki for the help with animation. This work has been supported by the Deutsche Forschungsgemeinschaft grants (SO1054/1 and SP2202/SO1054/2 to I.S., SPP 2202/LE721/17-1 to H.L. and SFB1064 to I.S. and H.L.) and National Institutes of Health grants (HG007743 to H.L., HG003143 to J.D. and GM114190 to L.M. by the Center for 3D Structure and Physics of the Genome of NIH 4DN Consortium, DK107980). J.D. is an investigator of the Howard Hughes Medical Institute. I.S. is deeply thankful to H. C. Macgregor for his guidance.

Author contributions

I.S. conceived the project. S. Leidescher, S.U., Y.F., K.T. and I.S. obtained biological samples. I.S., Y.F., S. Leidescher and S.U. conceived and performed microscopy and image analysis. S. Leidescher, C.M. and S. Link performed RNA-seq and ChIP–seq experiments. S.B. performed RNA-seq and ChIP–seq analyses. Y.F. and E.H. performed Hi-C experiments. E.H., J.D., J.R. (formerly known as J. Nübler), A.G. and L.M. performed Hi-C analysis. J.R. with contribution from L.M. performed simulations. S. Leidescher, S.U., J.R., A.G. and I.S. prepared the figures. I.S. wrote the manuscript with contributions from S. Leidescher, Y.F., S.U., J.R., K.T., E.H., H.L., J.D. and L.M.

Competing interests

The authors declare no competing interests.

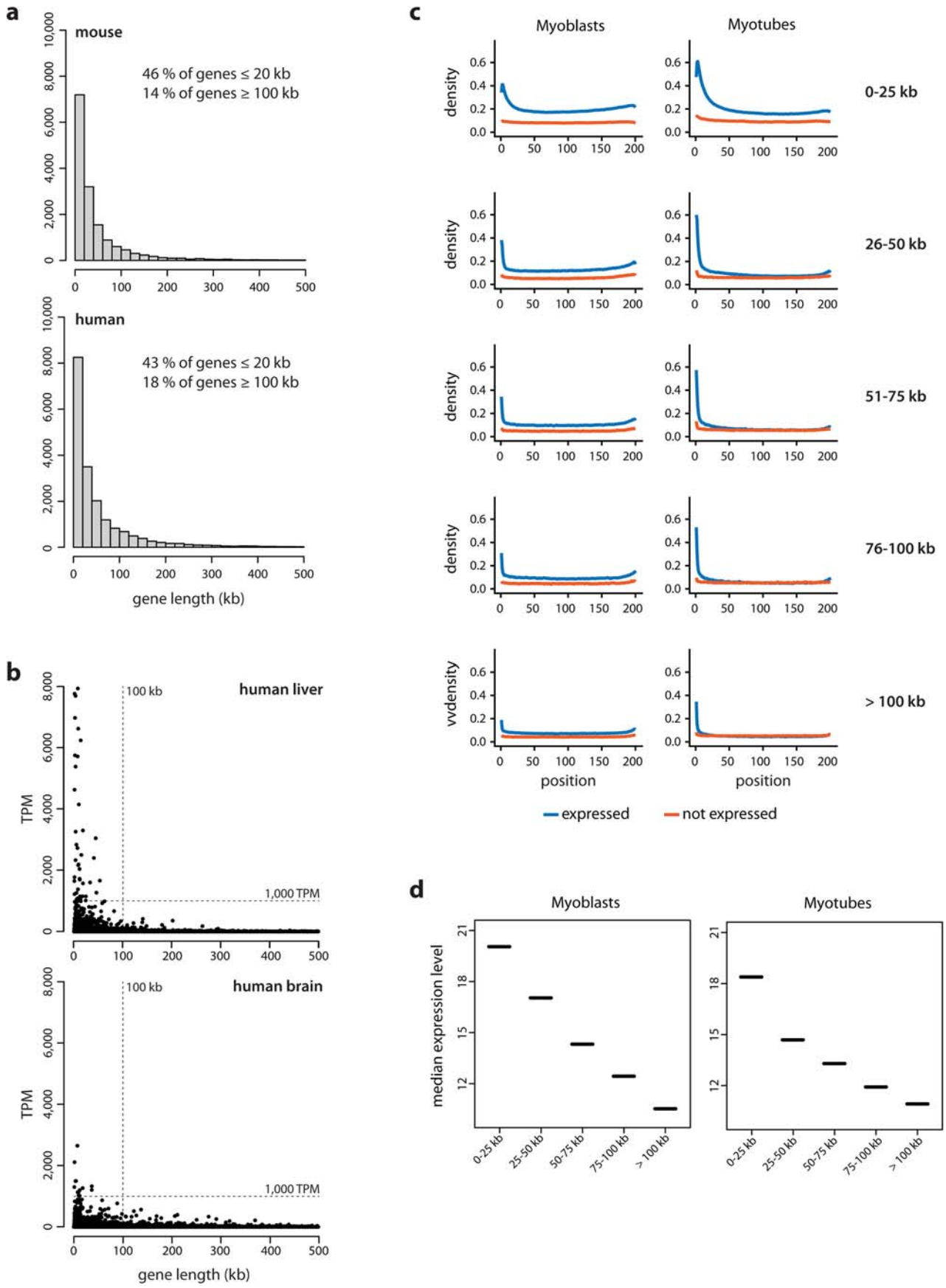
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41556-022-00847-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-022-00847-6.

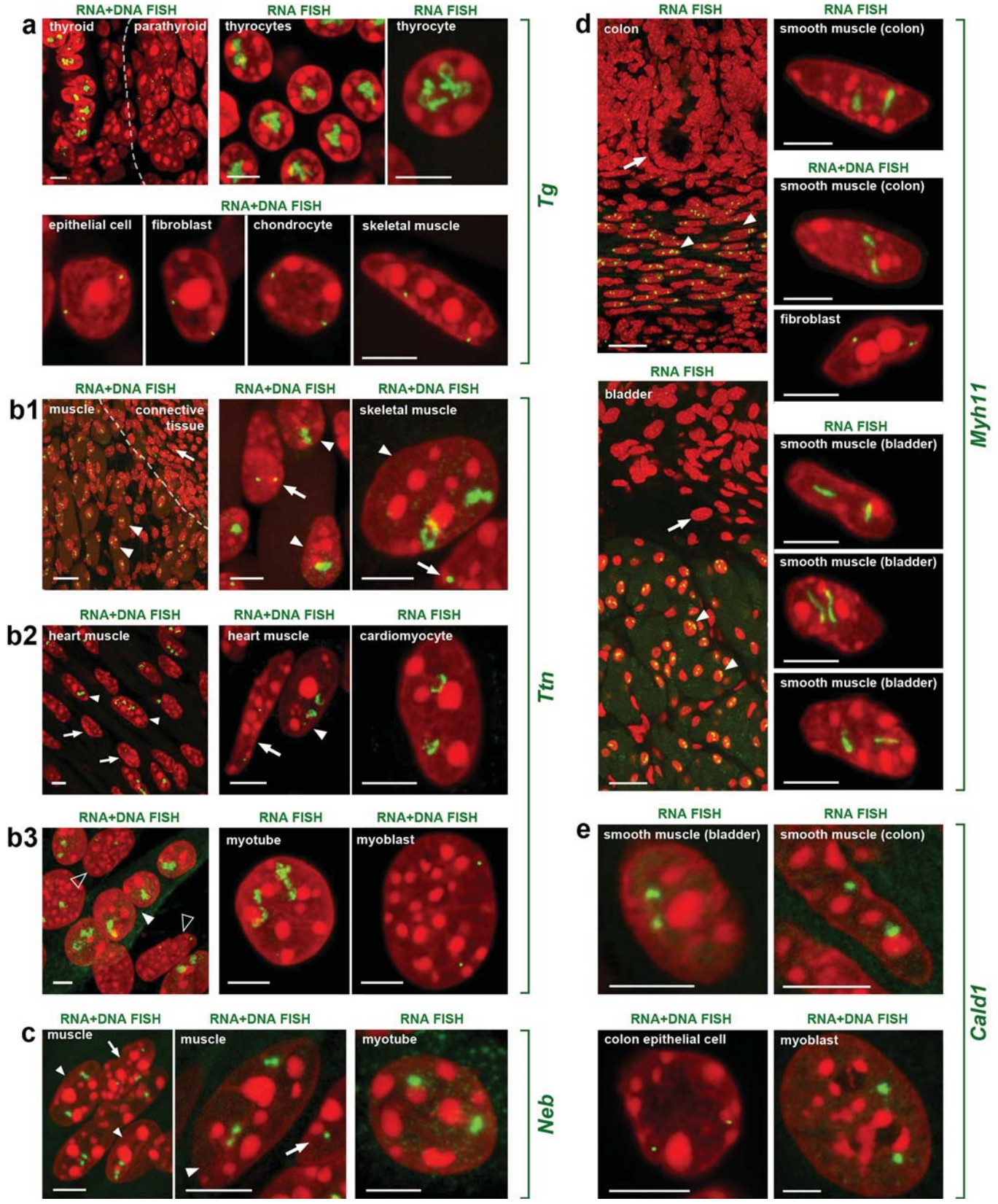
Correspondence and requests for materials should be addressed to Leonid Mirny or Irina Solovei.

Peer review information *Nature Cell Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.



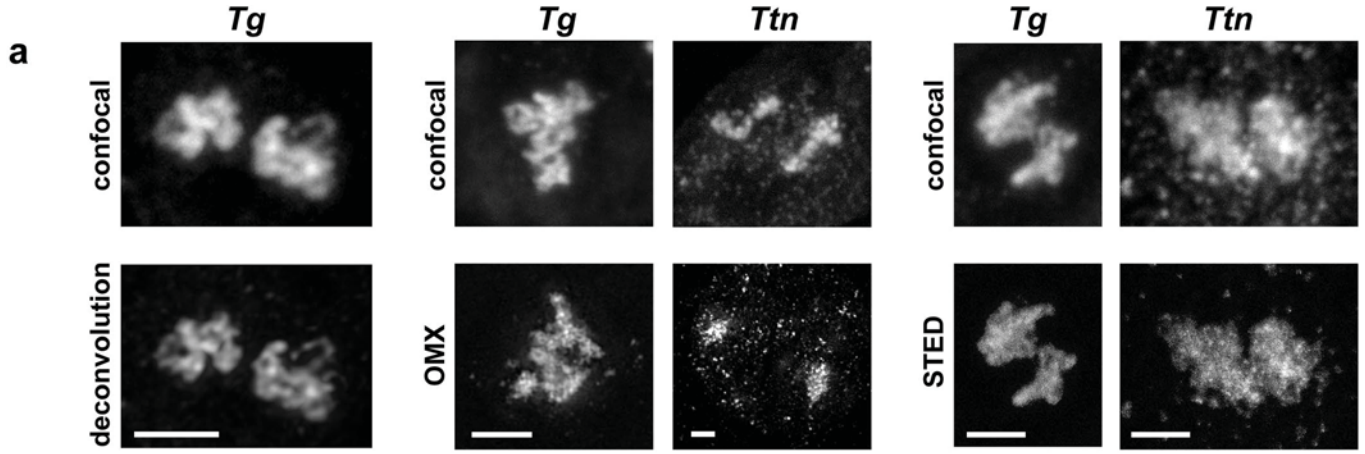
Extended Data Fig. 1 | See next page for caption.

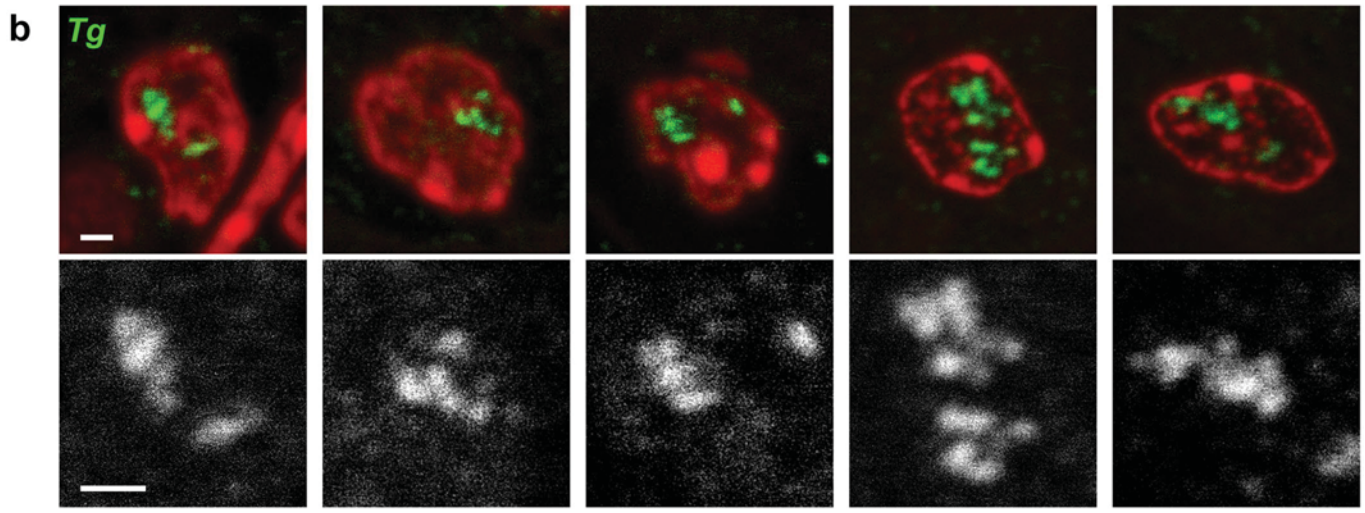
Extended Data Fig. 1 | Long genes are rare and expressed at lower levels than short genes. a, Analysis of gene length distribution within the human and mouse genomes showed that about 43% and 46% of all protein coding genes, respectively, have a length ≤20 kb and only 18% and 14% have a length of 100 kb or above. Bin size: 20 kb. Genes are annotated according to GENCODE. Only genes with a length <500 kb are shown. **b**, To select suitable genes for visualization with light microscopy, we studied gene expression profiles across 50 human tissues using the publicly available Genotype-Tissue Expression database (GTEx Consortium) and found that long genes, as a rule, are not highly expressed. For example, in liver (*top*) and brain (*bottom*) there were no expressed genes with both a length ≥100 kb and with a median expression ≥1,000 TPM. **c**, Comparison of RNAPII occupancy between short and long expressed genes. ChIP-seq with an antibody against the CTD of RNAPII in cultured mouse myoblasts (*left*) and *in vitro* differentiated myotubes (*right*). 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 the respective position within the gene (*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 (TPM) is higher in groups containing shorter genes (<25 kb) and generally negatively correlates with gene length.

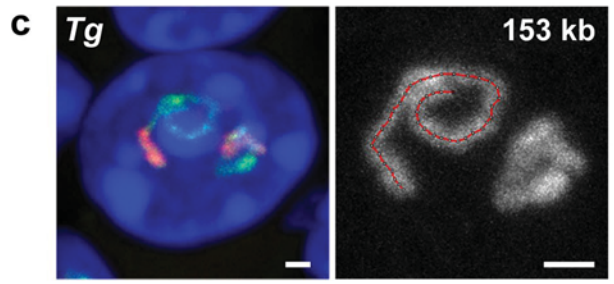


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | 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 expanding 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 Pmi28 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 where they form TLs. 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. As indicated above the panels, images display signals after either RNA-FISH (no tissue/cell DNA denaturation and no RNasing), or simultaneous detection of DNA and RNA (tissue/cell DNA denaturation but no RNasing). All images are *projections of 1–3 μm* confocal stacks. Scale bars for overviews of skeletal muscle, colon and bladder, *50 μm;* for the rest of the panels, *5 μm*. Data represent 100 in a,b3 and 10 in b1,b2,c-e independent experiments.

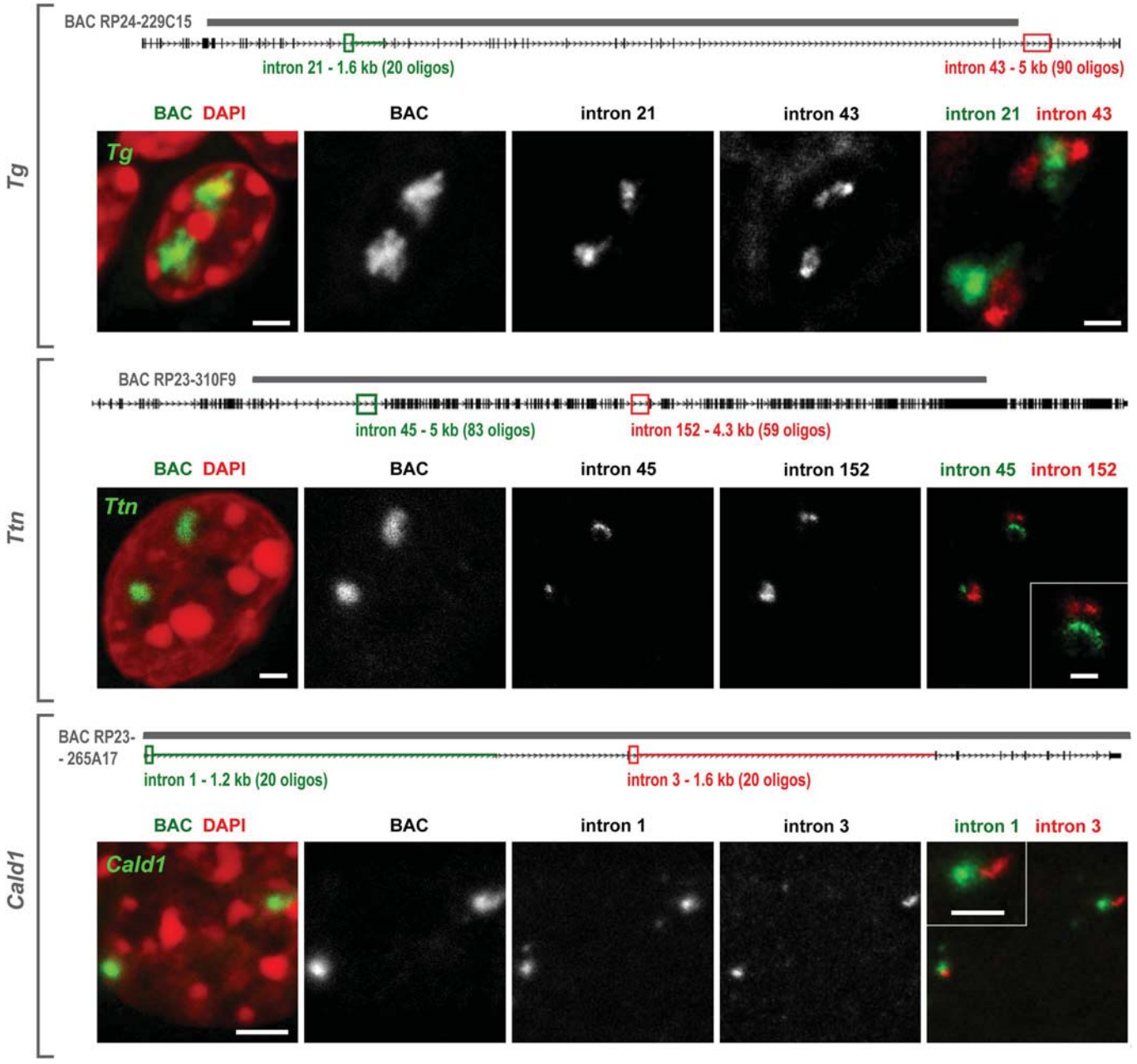




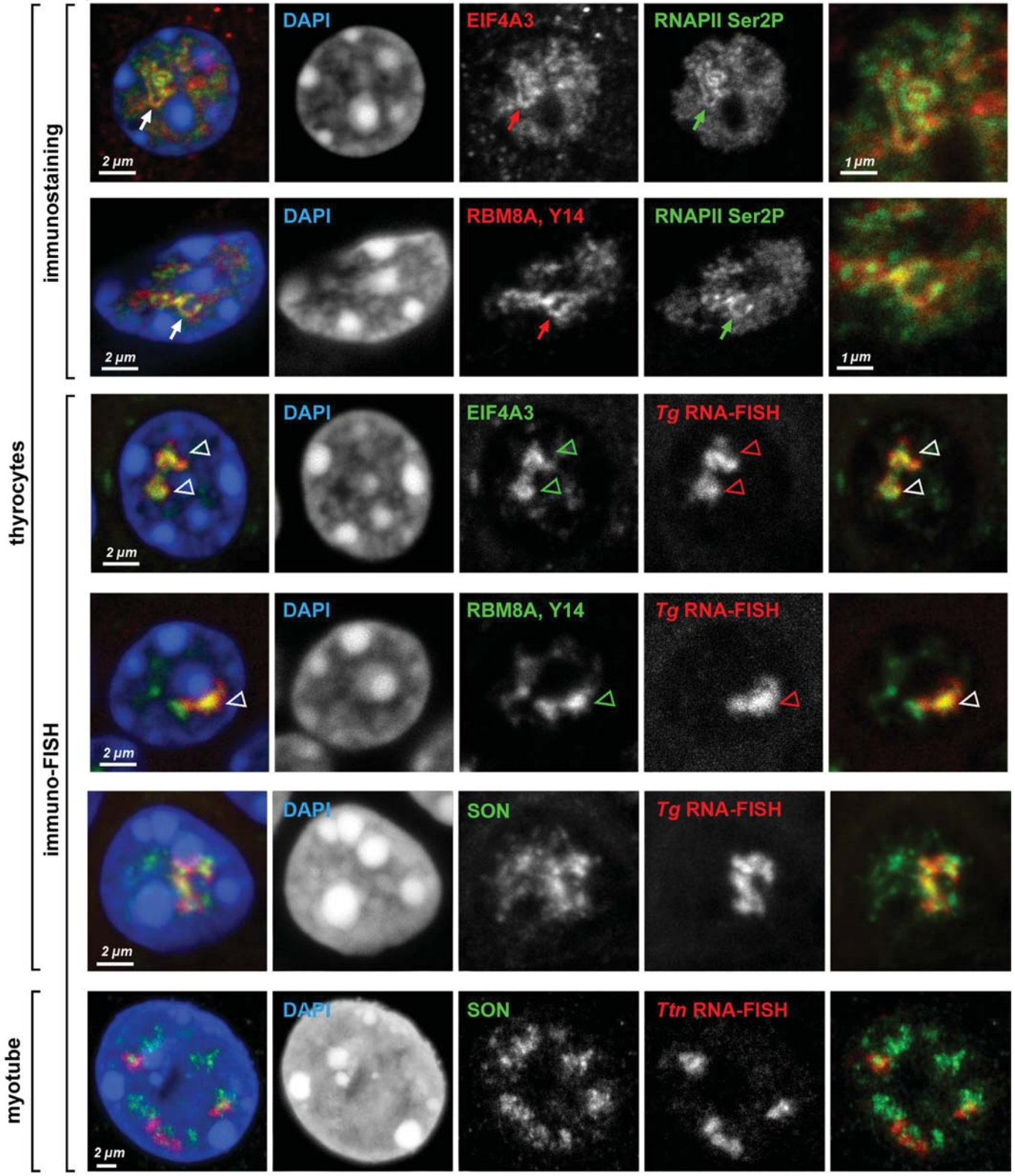


length of segments *(kb)* detected by FISH	153	109	62
calculated length of corresponding DNA *(µm)* without nucleosomes	52	37	21
calculated length of corresponding chromatin *(µm)* with nucleosomes	8	6	3
measured length of FISH signal *(µm)*	9	6	4
estimated compaction of labeled chromatin *(kb/µm)*	17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

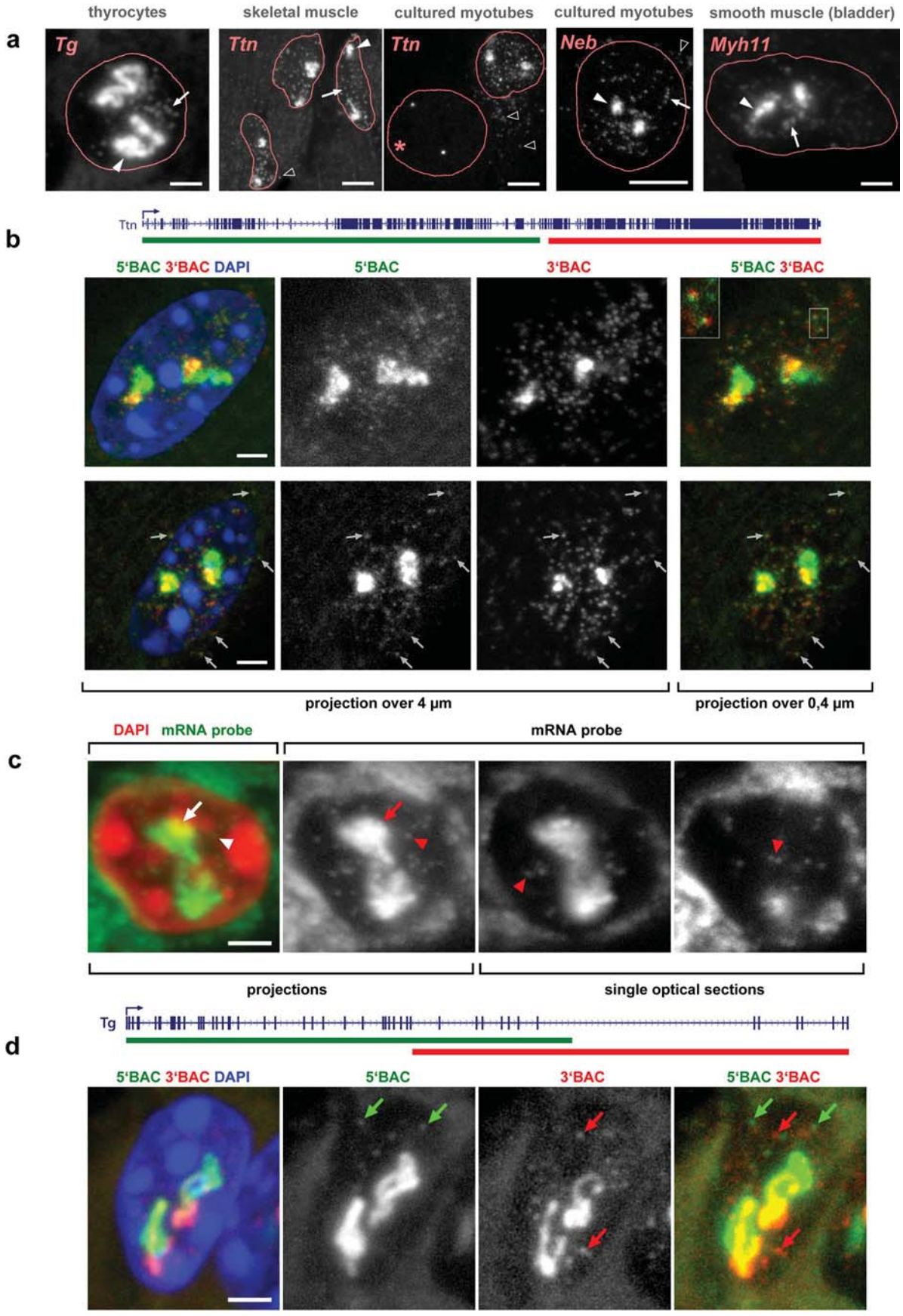
Extended Data Fig. 3 | Structure and compaction of TLs. a, The internal structure of *Tg* and *Ttn* TLs is not resolvable after deconvolution (*left*) and high resolution microscopy (*mid, right*). **b**, Coiling and folding of TLs demonstrated in 50–70 nm thin resin sections. The upper panel shows thin sections through nuclei of thyrocytes stained with DAPI (*red*) and *Tg* TLs detected by RNA-FISH (*green*). The lower panel shows 2-fold close-ups of the corresponding *Tg* TL as grey-scale images. Note curling and twisting of the loops. Images are single optical sections. **c**, To assess the compaction level of TLs, the contour length of three *Tg* TL regions was measured on projections after RNA-FISH. The track of the *Segmented Line* tool in *ImageJ*, used for measurements, is shown on the *right panel*. *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 correspond to a nucleosomal structure of chromatin (*table on the right*). However, since *Tg* TLs display internal structures and since the measurements were performed on maximum intensity projections, the compaction level of *Tg* TLs is probably overestimated. Scale bars: a, *2 µm*; b,c, *1 µm*. Data represent 3 independent experiments in a-c.



Extended Data Fig. 4 | TLs manifest co-transcriptional splicing. Two sequentially positioned introns were labeled with oligoprobes encompassing 1.2–5 kb. The schematics above the panels depict the distribution of oligoprobes (*green and red rectangles*), labeled introns (*green and red lines*) and positions of BAC probes used as references (*grey lines above genes*). After RNA-FISH the intron probes label TLs only partially and sequentially. Since the 5′ and 3′ intron signals do not overlap, the 5′ introns are spliced before the 3′ introns are read. For instance, in *Cald1*, introns 1 and 3 are separated by intron 2, suggesting that the "green" intron 1 is spliced out before polymerases reach the "red" intron 3, most likely after RNAPII runs over the 3′ splice-site of the first intron. Projections of confocal sections through 2–3.5 μm. Scale bars: 2 μm, in close-ups, 1 μm. Data represent 2 independent experiments.

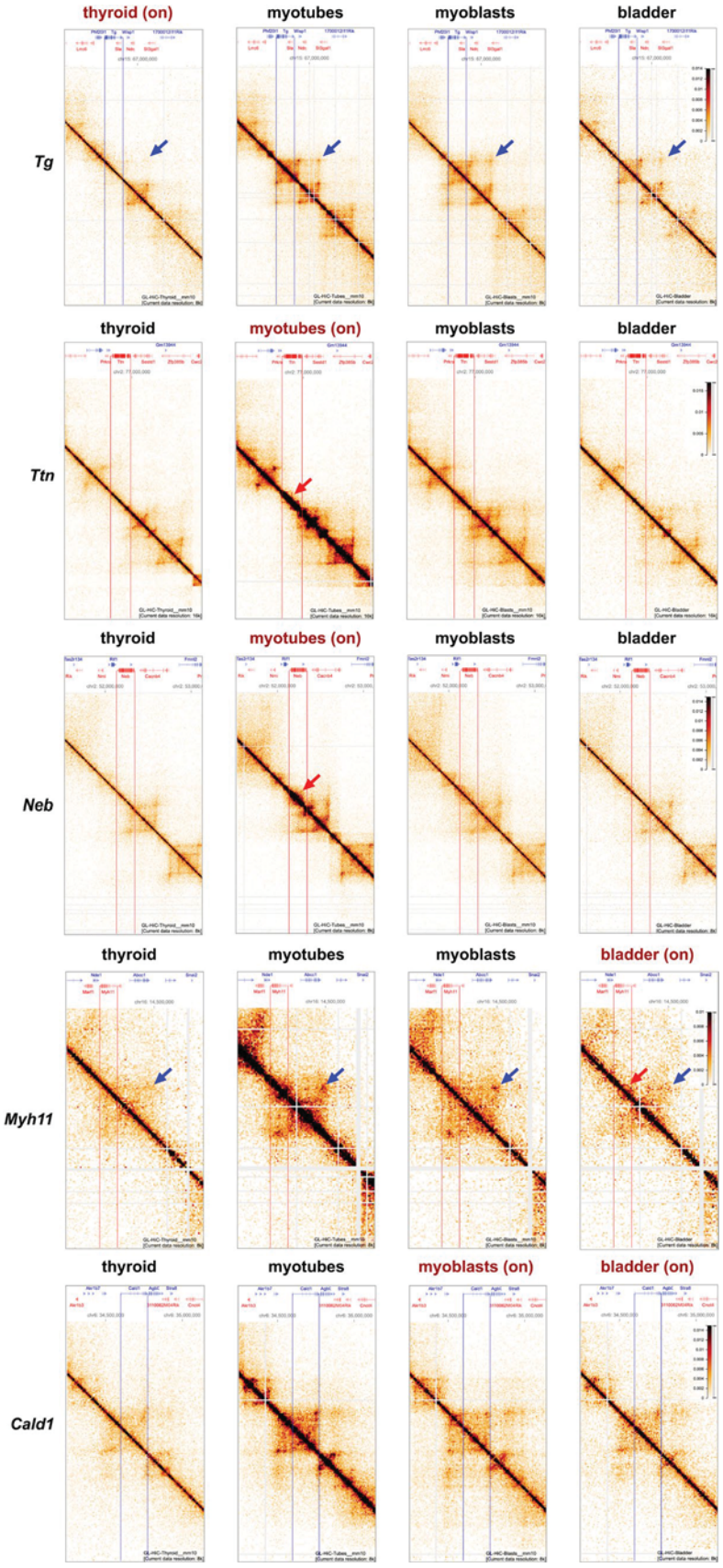


Extended Data Fig. 5 | Close association of TLs with splicing factors. Splicing factors (SON) and components of exon-junction complexes EIF4A3 and RBM8A (Y14), are either co-stained with RNAPII Ser2P (*the two top rows*) or visualized together with TLs in immuno-FISH (*the rest of the rows*). Note that signals of TLs and splicing factors colocalize only partly. The myotube nucleus is tetraploid and thus exhibits 4 *Ttn* RNA signals. Images are partial projections of either 0.6 *μm* (for immunostaining) or 0.9 *μm* (for immuno-FISH). Scale bars: *2 μm*, in close-ups, *1 μm*. Data represent 3 independent experiments.

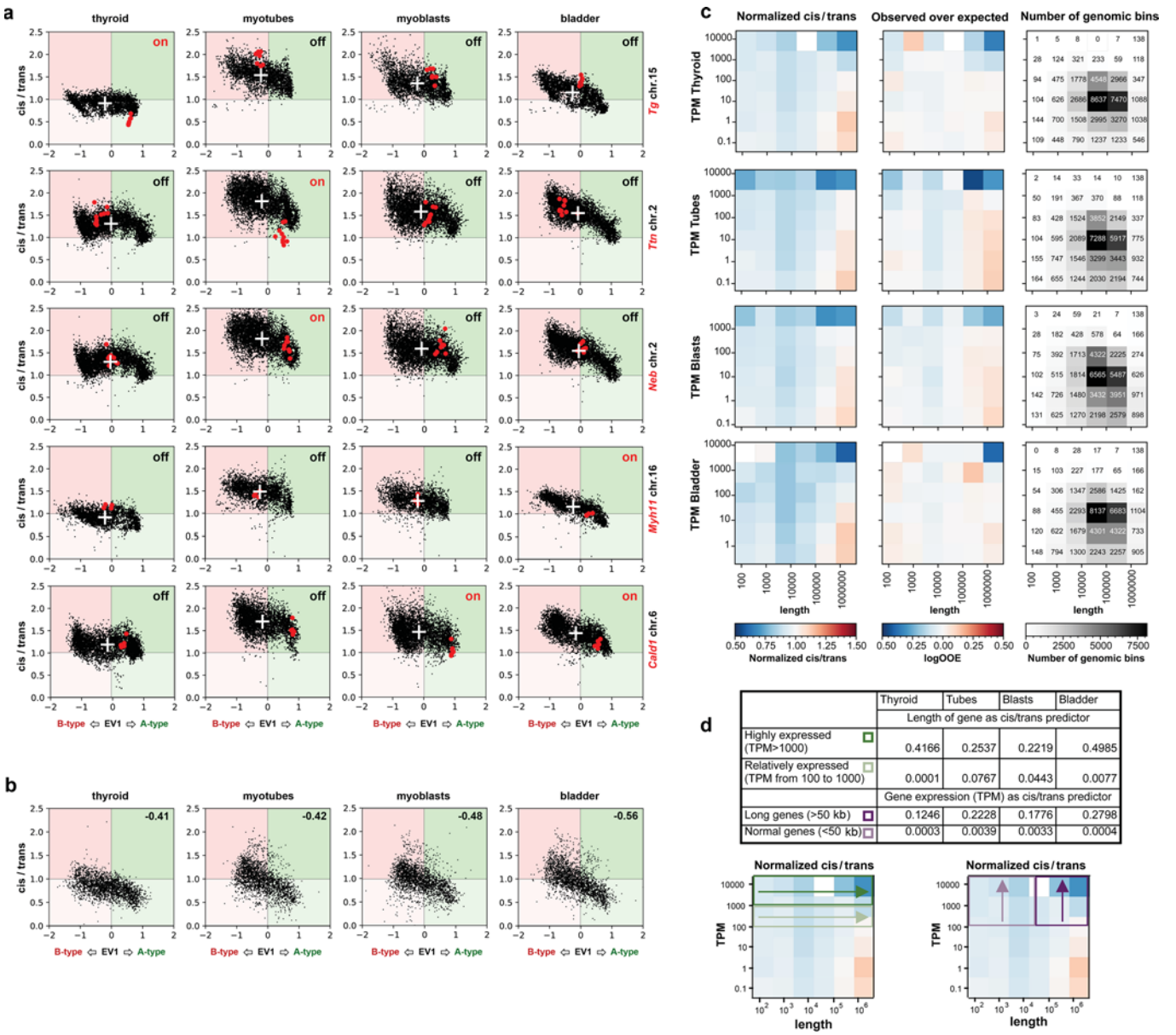


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Nucleoplasmic granules in cells with highly expressed long genes. a, RNA-FISH reveals numerous nucleoplasmic granules (*arrows*) surrounding TLs (*arrowheads*) after hybridization with genomic probes. For clarity, only RNA signals are shown within the outlined nuclei. Empty arrowheads point at similar granules in the myotube cytoplasm. The asterisk marks the nucleus of a myoblast not expressing *Ttn*. **b**, In muscles and cultured myotubes, the majority of granules (81%) are double-labeled with probes for the 5′ and 3′ halves of *Ttn* and found in both the nucleoplasm and cytoplasm (*arrows on the lower panel*), thus likely representing *Ttn* mRNAs. Remarkably, the 5′ and 3′ signals are spatially distinguished within the granules (*insertion*) presumably due to the exceptionally long *Ttn* mRNA of ca 102 kb. The observed separation of the 5′ and 3′ halves of *Ttn* mRNA is in agreement with previously described structures of cytoplasmic mRNPs[28–30]. **c**, In difference to *Ttn*, mRNAs of *Tg, Neb, Cald1* and *Myh11* genes are short (4–20 kb) and can be only detected with oligoprobes specifically hybridizing to all exons. Thus the oligoprobe for all 48 *Tg* exons hybridizes to nRNAs decorating TLs (*arrows*) and also labels multiple nucleoplasmic granules (*arrowheads*). **d**, The majority of the other thyrocyte nucleoplasmic granules are labeled with either 5′ (*green*) or 3′ (*red*) genomic probes with only 10% of granules being double-labeled. The brightness of the RNA-signal on the *most right panel* is purposely increased to highlight nucleoplasmic granules (*green and red arrows*). Such differential labeling of nucleoplasmic granules, exemplified here for thyrocytes, is characteristic for other studied genes and strongly suggests that these granules represent accumulations of excised introns. The distribution of the used BAC probes in respect to the studied genes are depicted above the image panels. Scale bars: a, *2 μm* for *Tg* and *Myh11*, *5 μm* for *Ttn* and *Neb*; b, c, d, *2 μm*. Data represent 10 in a,b and 3 in c,d independent experiments.

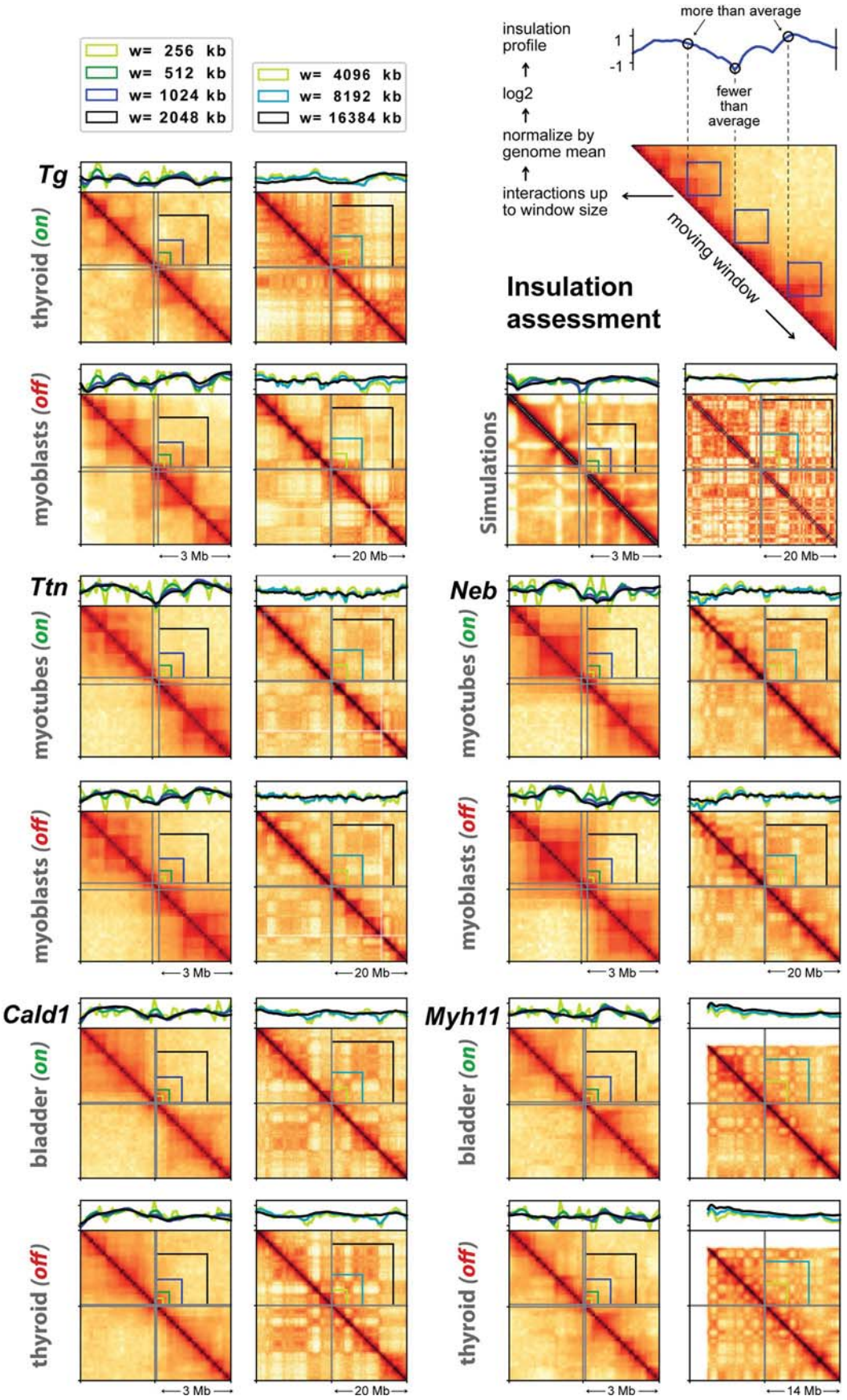


Extended Data Fig. 7 | Fragments of Hi-C maps around TL-forming genes in corresponding tissues and cells. Each row shows the same region around indicated genes in different tissues or cells; *"on"* indicates a cell type where a gene is active, allowing to compare changes associated with TL formation. *Blue arrows* indicate loss of TAD borders and associated dots of CTCF-CTCF enrichment for expressed *Tg* and *Myh11*. *Red arrows* indicate an increase in self-interactions within genes visible for expressed *Ttn, Neb* and *Myh11*.



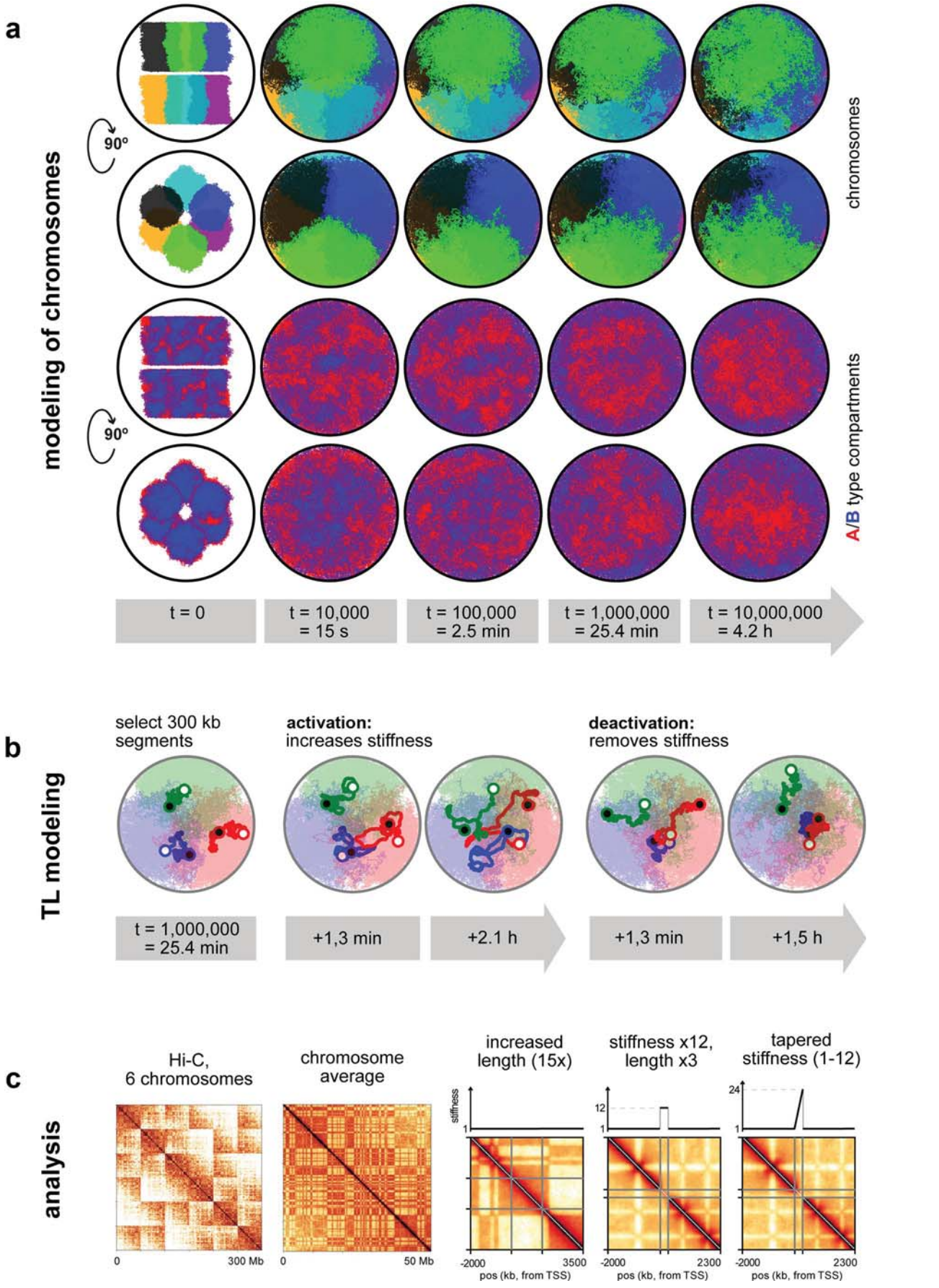
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | *Cis-to-trans* contact ratios and A/B compartments in studied cells. a, *Cis-to-trans* ratios and compartment affiliations for 5 studied genes. The scatter plots are computed from the compartment profiles and the *cis-to-trans* ratio profiles of the chromosomes harboring the genes at a bin size of 32 kb. Genes of interest are highlighted with red dots; the white crosses mark the chromosome means of the compartment and *cis-to-trans* ratio profiles. *Tg, Ttn, Neb* and *Myh11* move from B to A compartment upon their activation; *Cald1* is found in A compartment not only in myoblasts and smooth muscle, as expected, but also in thyroid and myotubes. This is in agreement with the low *Cald1* expression also in thyroid samples enriched in blood vessels and in samples of myotubes that normally include up to 20% of myoblasts, as well as with localization of *Cald1* in a gene-dense region of chromosome 6. Note that the *Ttn* and *Neb* genes tend to be in A compartment, although to a lesser degree, also in myoblasts, which can be explained by the presence of cells that started their differentiation into single-cell myotubes. **b**, *Cis-to-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-to-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. **c**, Externalization of the expressed genes from their harboring chromosomes measured by the *cis-to-trans* ratios as a function of gene length and expression in corresponding tissues. *Left column of heatmaps*: median *cis-to-trans* normalized by chromosome. Notice the reduction of *cis-to-trans* (that is, increasing externalization) for highly transcribed genes (top rows in each heatmap) as gene length increases (moving from left to right). Similarly, *cis-to-trans* goes down for long genes (*right-most column* of each heatmap) as expression increases (going up along this column). Interestingly, lowly transcribed long genes (*lower right corner*) have high *cis-to-trans*, indicating strong internalization, but become strongly externalized as they become highly expressed (*upper right corner*). *Middle column*: median *cis-to-trans* ratios controlled for compartmental signal (the first eigenvector, E1). Heatmaps show the logarithm of observed median *cis-to-trans* ratio divided by the expected given E1 in the corresponding bin. For the expected value, all the genomic bins were separated into 20 ranges by their E1, and the median *cis-to-trans* for each range was considered as expected. Notice that for most genes, E1 explains most of the *cis-to-trans* ratio. However, *cis-to-trans* is considerably lower for extremely long and extremely highly transcribed genes (*upper right corner* of each heatmap). *Right column*: the number of genomic bins in each range of length and transcription. Notice that very few genes show high externalization. **d**, *Table* of coefficients of determination (R2) for regression of *cis-to-trans* ratios of the genomic bins normalized by chromosome, in four tissues. Only the bins of expressed genes (TPM > 100) are considered. Gene length is an excellent predictor of *cis-to-trans* ratio for genomic bins of highly expressed genes (TPM > 1000, *1st row*), but not for other expressed genes (TPM from 100 to 1000, *2nd row*). Gene expression is a good predictor of *cis-to-trans* ratio for genomic bins of long genes (TPM > 50 kb, 3rd row), but not for shorter genes (TPM < 50 kb, last row). *Heatmaps*: visual illustration of gene subsets in this analysis.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | TLs do not cause insulation at different length scales. Insulation assesses Hi-C contacts spanning across a given locus up to a maximal distance *w* (*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 a 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 non-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 ≈ 1 Mb in both thyroids (*on*) and myoblasts (*off*), and no dip in either cell type on the larger scale. Analysis of simulated TLs (Fig. 8 and Extended data Fig. 10) confirmed that TL formation does not cause large scale insulation (*bottom row*).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | 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, the 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**, TLs are modeled by choosing a 300 kb segment on each chromosome 25.4 minutes after expansion and increasing the stiffness of the polymer fiber. The genes quickly expand on the order of minutes and are simulated for approximately 1.5 h. The measurements of inter-flank distances and Hi-C maps are 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 of a typical segregation of A- and B-type chromatin. The three rightmost graphs show zoomed views of modeled genes with stiffness profiles above the maps.

nature portfolio

https://doi.org/10.1038/s41556-022-00847-6

Supplementary information

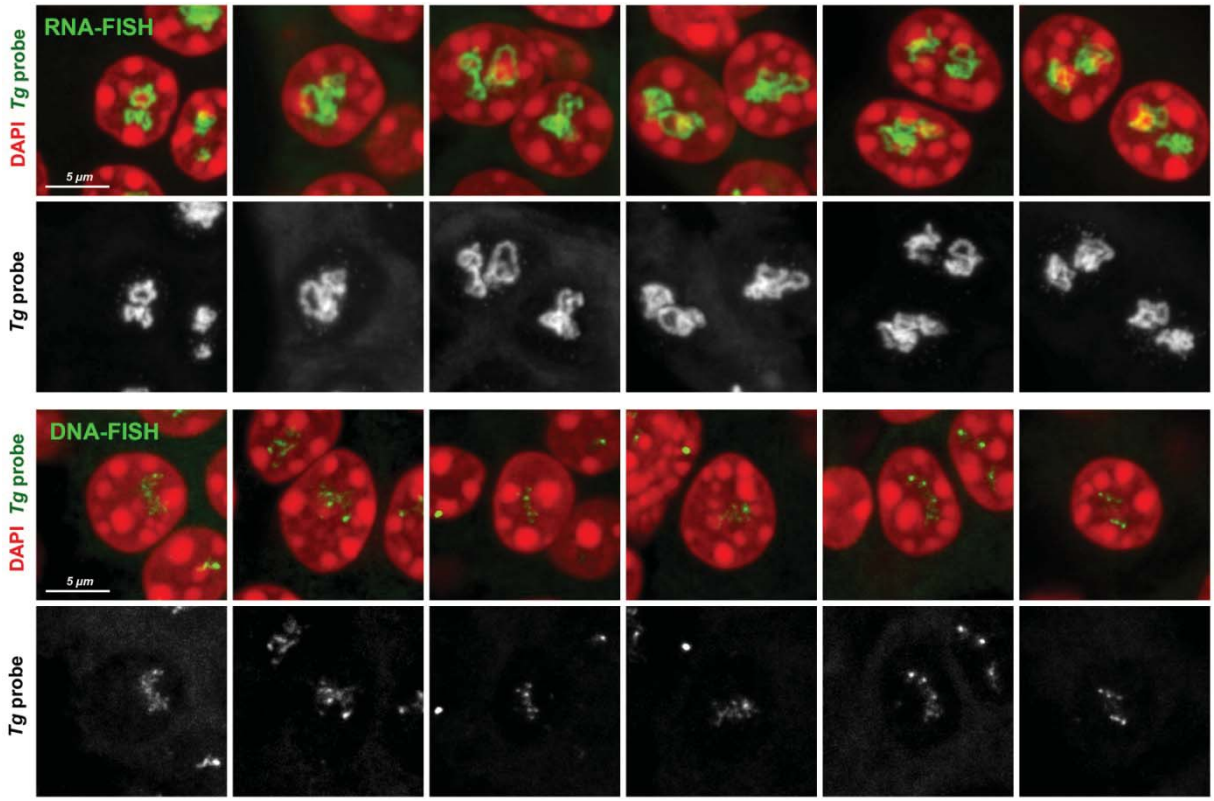
Spatial organization of transcribed eukaryotic genes

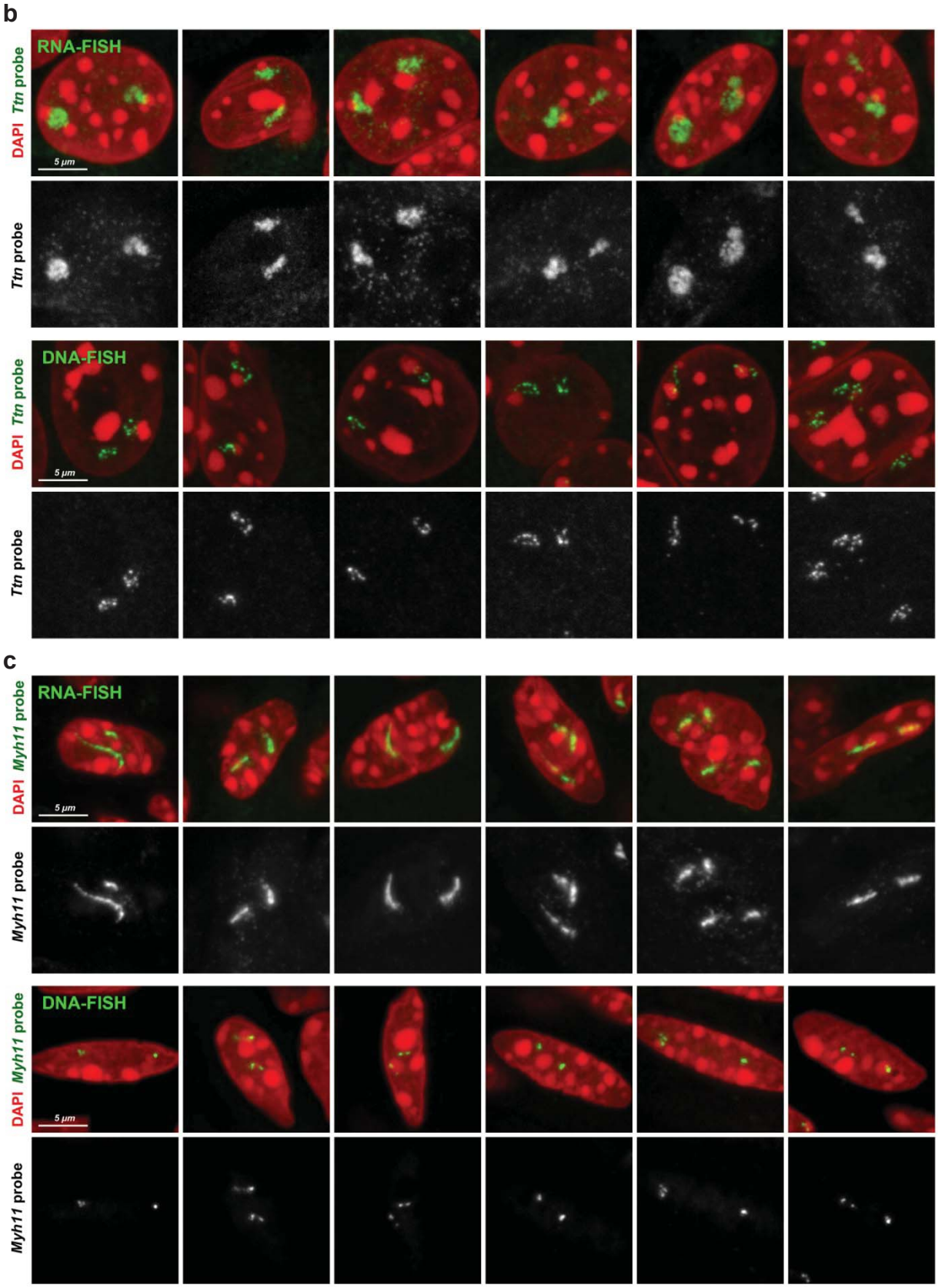
In the format provided by the authors and unedited

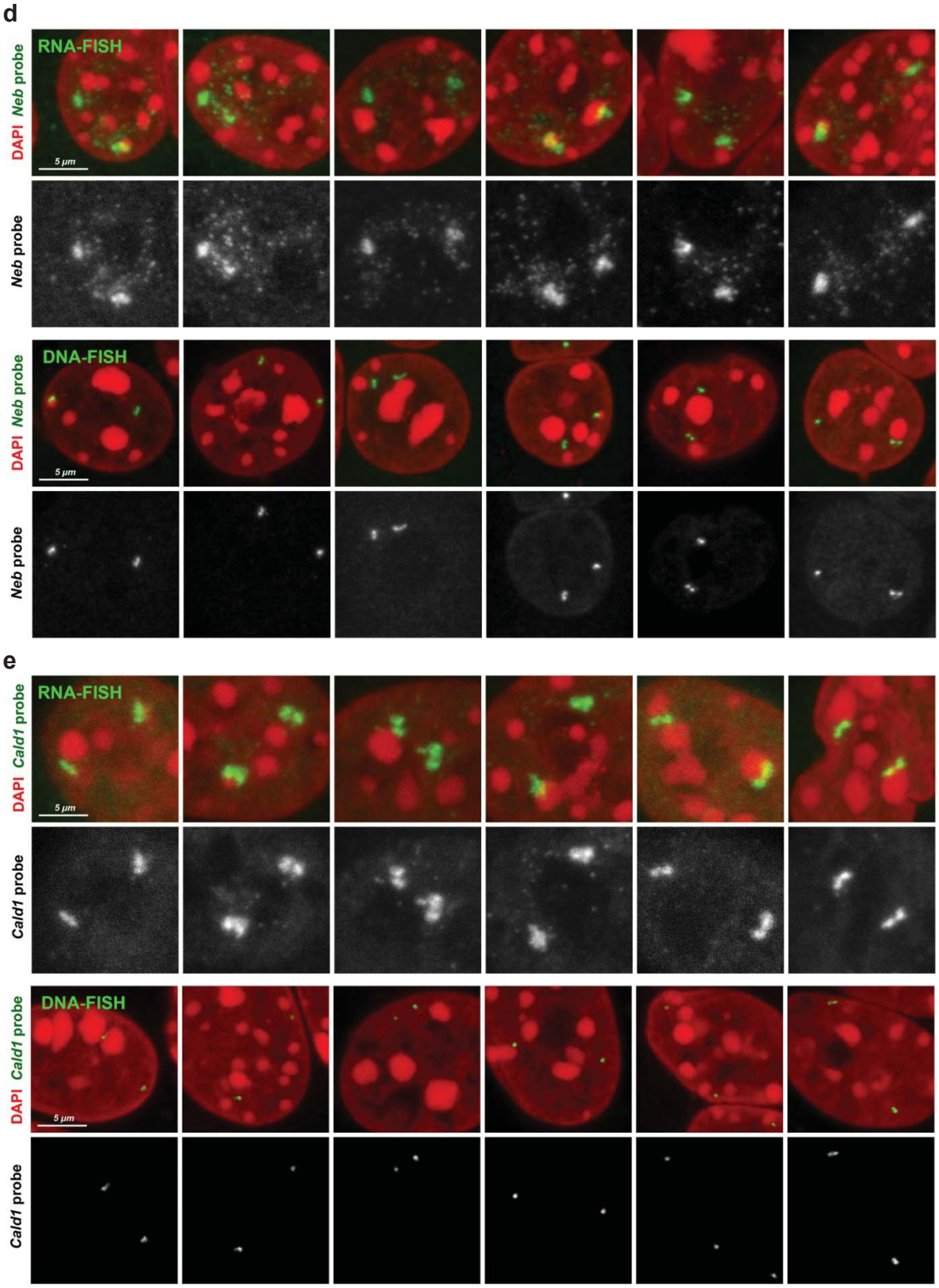
Supplementary Information Figures

Supplementary Information Figure 1. Exemplary images of RNA- and DNA-FISH signals for the Tg gene in thyroid sections (**a**), for the Ttn gene in cultured myotubes (**b**), for the Myh11 gene in bladder sections (**c**), for the Neb gene in cultured myotubes (**d**) and for the Cald1 gene in cultured myoblasts (**e**).

a

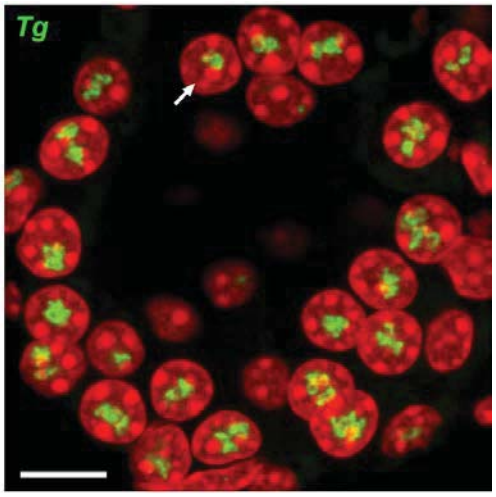



b
RNA-FISH
DAPI Ttn probe
5 μm
Ttn probe
DNA-FISH
DAPI Ttn probe
5 μm
Ttn probe
c
RNA-FISH
DAPI Myh11 probe
5 μm
Myh11 probe
DNA-FISH
DAPI Myh11 probe
5 μm
Myh11 probe

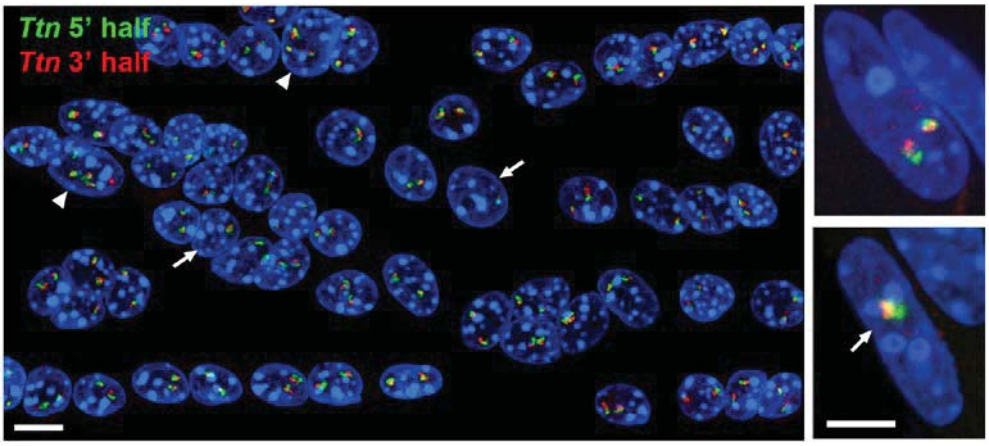

d
RNA-FISH
DAPI Neb probe
5 μm
Neb probe
DNA-FISH
DAPI Neb probe
5 μm
Neb probe
e
RNA-FISH
DAPI Cald1 probe
5 μm
Cald1 probe
DNA-FISH
DAPI Cald1 probe
5 μm
Cald1 probe

Supplementary Information Figure 2. Exemplary images of RNA-FISH signals for the *Tg* gene in thyrocytes in thyroid sections (**a**), for the *Ttn* gene in cultured myotubes (**b**) and for the *Myh11* gene in smooth muscles in bladder sections (**c**). All images are partial projections (5-10 μm) of confocal stacks. **a**, One thyrocyte in the entire stack (20 μm) exhibits one transcriptionally active allele (arrow), scale bar, 10 μm. **b**, Projections through nuclei of cultured myotubes (left) and skeletal muscle (right) after RNA-FISH with probes detecting 5' and 3' *Ttn* nRNAs. Arrows point at the nuclei with only one allele active; arrowheads point at the two tetraploid nuclei. Scale bar for cultured myotubes, 10 μm; for skeletal muscle, 5 μm. **c**, Projections through colon smooth muscle after RNA-FISH with the probe for *Myh11*. Smooth muscle cells underlay colon epithelium not expressing *Myh11* (asterisks mark intestinal crypts). One of the nuclei in the closeup on the right exhibits only one expressing allele (arrow). Scale bar on the left, 50 μm; on the right, 5 μm

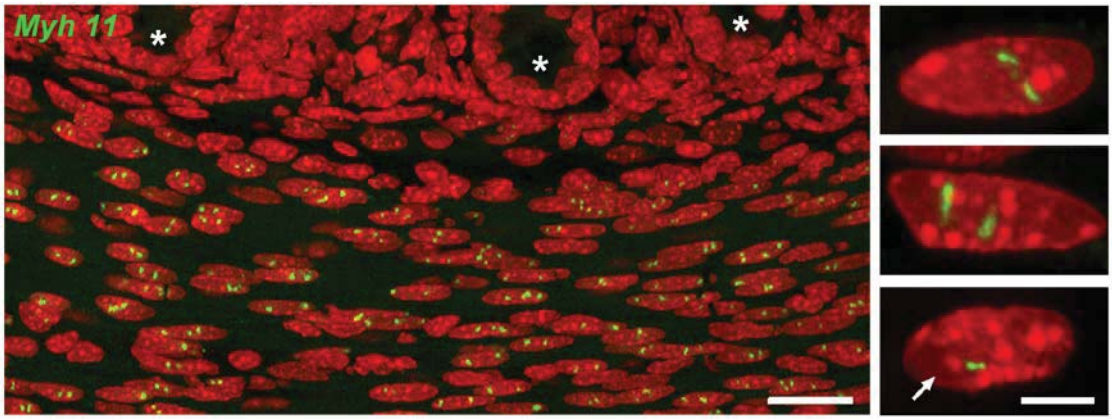
a



b



c

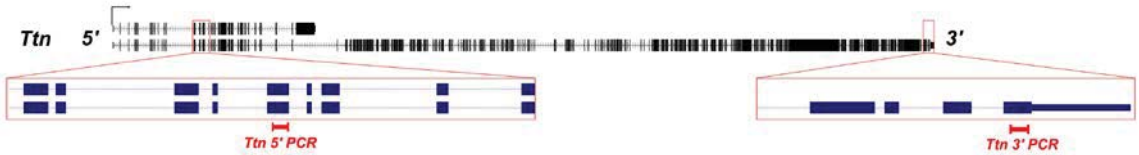


Supplementary Figure 3. Activation of the *Ttn* gene with dCas9-VPR.

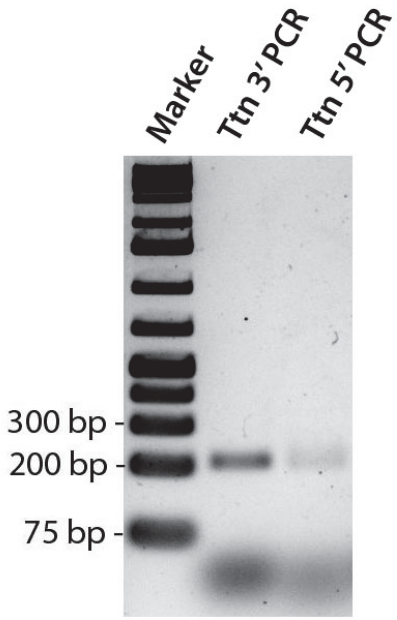
a, Localization of primers used to estimate level of *Ttn* expression after activation with dCas9-VPR. **b**, Localization of primers used to verify expression of the full length Ttn after activation with dCas9-VPR. **c**, Fragments amplified by PCR with primers for 5' and 3' ends of *Ttn* mRNA.

a

b

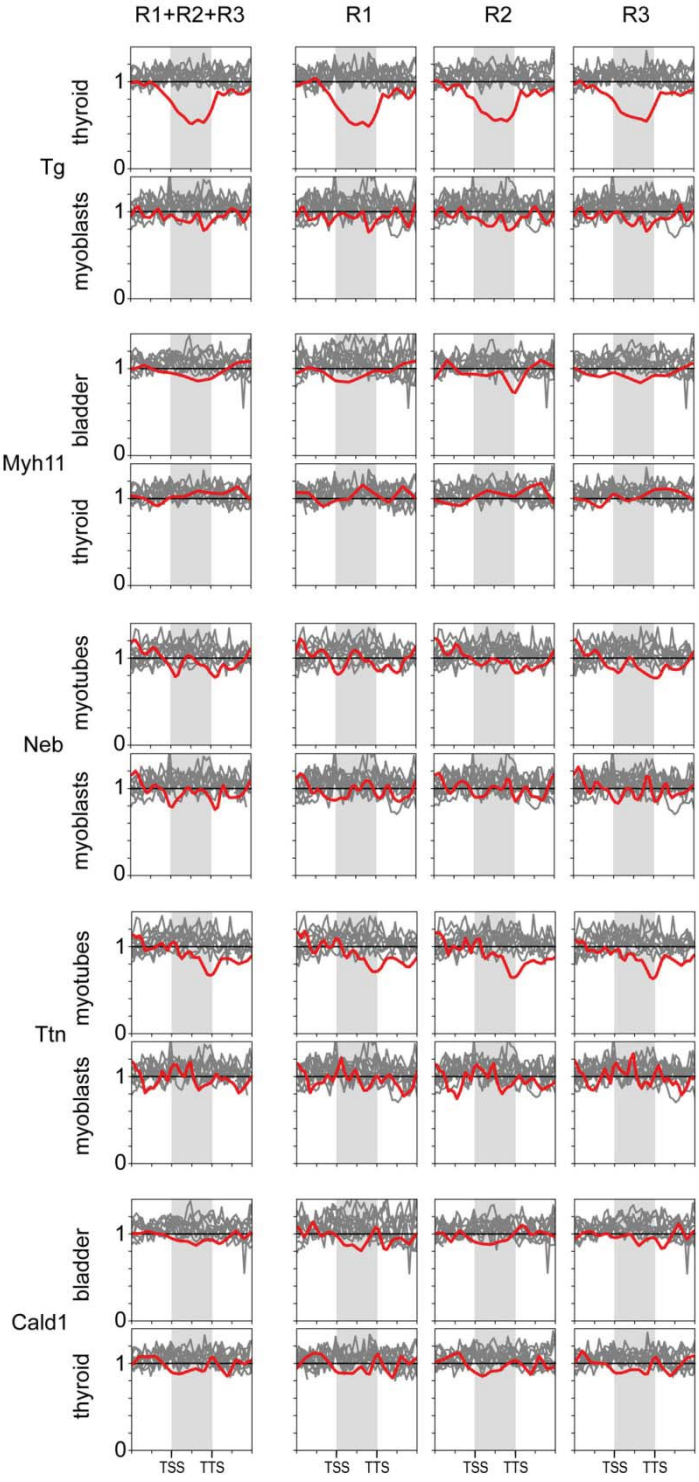


c



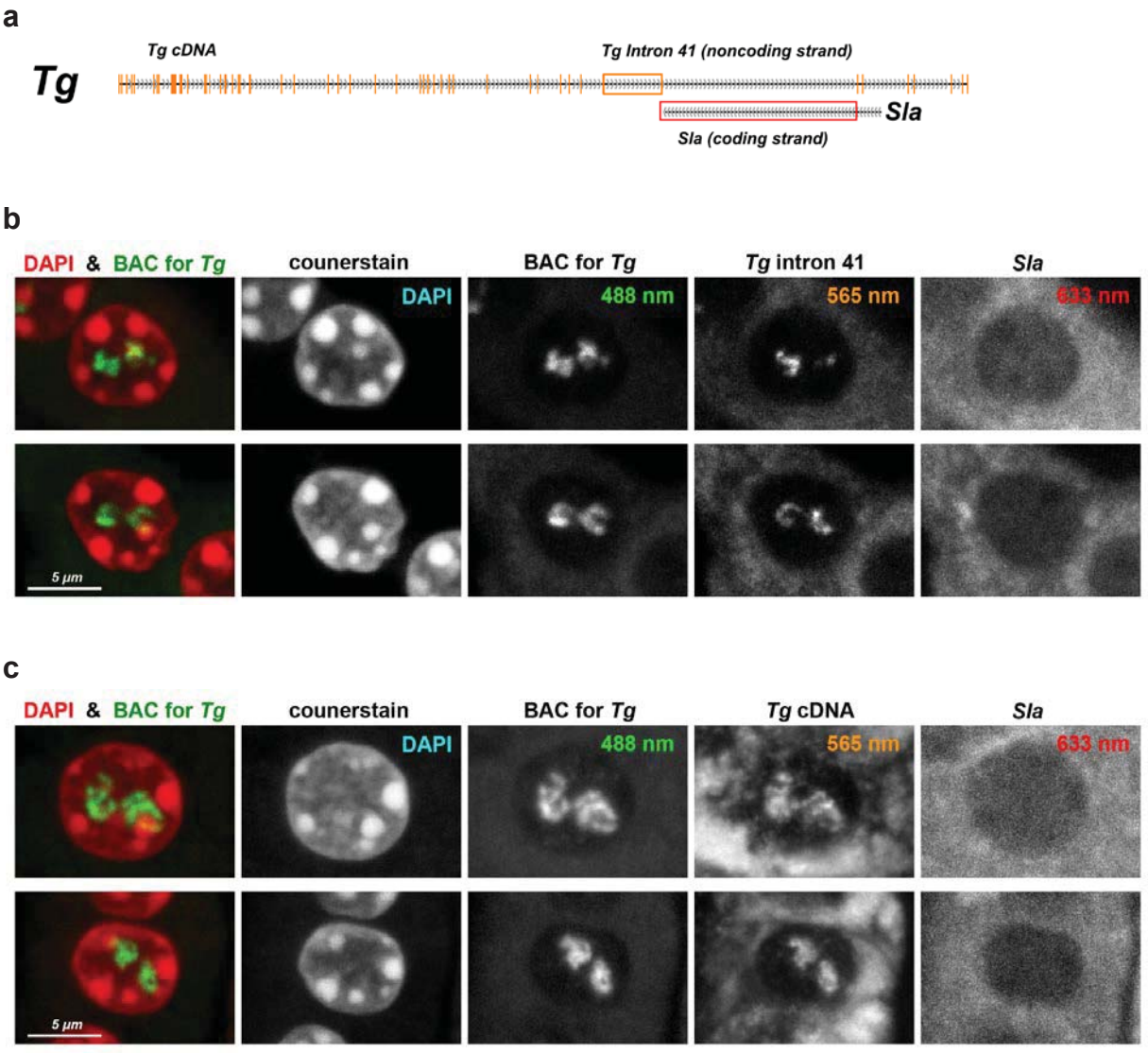
Supplementary Figure 4. Reproducibility of 3 replicates of Hi-C.

To determine the reproducibility of the replicates, we created *cis/trans* ratio profiles for each replicate analogous to those shown in Figure 6f, where the total number of Hi-C contacts of a locus with loci on the same chromosome is divided by the total number of contacts with other chromosomes, calculated near the gene of interest (*red*) and compared to 10 other long genes with low expression (*grey*). More details on how the plots were generated are given in the legend to Figure 6 and Methods section. See Supplementary Table 2 for the lists of genes



Supplementary Figure 5. The *Sla* gene is silent in thyrocytes. The strand opposite to encoding *Tg* intron 41 includes the *Sla* gene. The gene codes for an adapter protein negatively regulating T-cell receptor signaling and is expressed in blood cells. Although thyroid RNA-seq indicated *Sla* transcription at 240 TPM (Supplementary Table 2), RNA-FISH did not reveal *Sla* transcription in thyrocytes and thus we conclude that *Sla* expression does not contribute to the intron 41 extension and that *Sla* transcripts are present in thyroid samples due to abundant blood vessels. Therefore, we conclude that the strong expansion of the intron 41 (Figure 7d) is due to its own transcription.

a, Localization of oligo-probes used in RNA-FISH experiments for detection of *Sla* expression. **b**, Examples of thyrocyte nuclei after RNA-FISH with SABER oligoprobe for *Tg* intron 41 and SABER oligoprobe for nRNAs of the *Sla* gene. The BAC probe for *Tg* was used as a reference. **c**, Examples of thyrocyte nuclei after RNA-FISH with SABER oligoprobe for all *Tg* exons and SABER oligoprobe for nRNAs of the *Sla* gene. The BAC probe for *Tg* was used as a reference. Note, that brightness of the *Sla* channel (633 nm) is artificially enhanced in comparison to the other channels.



The highly and perpetually upregulated thyroglobulin gene is a hallmark of functional thyrocytes

Simon Ullrich, Susanne Leidescher, Yana Feodorova, Katharina Thanisch, Jean-Baptiste Fini, Bernd Kaspers, Frank Weber, Boyka Markova, Dagmar Führer, Mirian Romitti, Stefan Krebs, Helmut Blum, Heinrich Leonhardt, Sabine Costagliola, Heike Heuer, Irina Solovei

Frontiers in Cell And Developmental Biology, 2023 - https://doi.org/10.3389/fcell.2023.1265407

frontiers | Frontiers in Cell and Developmental Biology

TYPE Original Research
PUBLISHED 04 October 2023
DOI 10.3389/fcell.2023.1265407

OPEN ACCESS

EDITED BY
Kundan Sengupta,
Indian Institute of Science Education and Research, India

REVIEWED BY
Igor I. Kireev,
Lomonosov Moscow State University, Russia
Kaustubh Wagh,
National Institutes of Health (NIH), United States

*CORRESPONDENCE
Irina Solovei,
Irina.Solovei@lrz.uni-muenchen.de

†PRESENT ADDRESSES
Yana Feodorova,
Department of Medical Biology, Medical University of Plovdiv, Division of Molecular and Regenerative Medicine, Research Institute at Medical University of Plovdiv, Plovdiv, Bulgaria
Katharina Thanisch,
Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany

RECEIVED 22 July 2023
ACCEPTED 22 September 2023
PUBLISHED 04 October 2023

CITATION
Ullrich S, Leidescher S, Feodorova Y, Thanisch K, Fini J-B, Kaspers B, Weber F, Markova B, Führer D, Romitti M, Krebs S, Blum H, Leonhardt H, Costagliola S, Heuer H and Solovei I (2023), The highly and perpetually upregulated thyroglobulin gene is a hallmark of functional thyrocytes.
Front. Cell Dev. Biol. 11:1265407.
doi: 10.3389/fcell.2023.1265407

The highly and perpetually upregulated thyroglobulin gene is a hallmark of functional thyrocytes

Simon Ullrich[1], Susanne Leidescher[1], Yana Feodorova[1†], Katharina Thanisch[1†], Jean-Baptiste Fini[2], Bernd Kaspers[3], Frank Weber[4], Boyka Markova[5], Dagmar Führer[5], Mirian Romitti[6], Stefan Krebs[7], Helmut Blum[7], Heinrich Leonhardt[1], Sabine Costagliola[6], Heike Heuer[5] and Irina Solovei[1]*

[1]Biocenter, Ludwig Maximilians University Munich, Munich, Germany, [2]Département Adaptations du Vivant (AVIV), Physiologie Moléculaire et Adaptation (PhyMA UMR 7221 CNRS), Muséum National d'Histoire Naturelle, CNRS, CP 32, Paris, France, [3]Department for Veterinary Sciences, Ludwig Maximilians University Munich, Planegg, Germany, [4]Department of General, Visceral and Transplantation Surgery, Section of Endocrine Surgery, University Duisburg-Essen, University Hospital Essen, Essen, Germany, [5]Department of Endocrinology, Diabetes and Metabolism, University Duisburg-Essen, University Hospital Essen, Essen, Germany, [6]IRIBHM ULB, Brussels, Belgium, [7]Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, Ludwig Maximilians University Munich, Munich, Germany

Abnormalities are indispensable for studying normal biological processes and mechanisms. In the present work, we draw attention to the remarkable phenomenon of a perpetually and robustly upregulated gene, the thyroglobulin gene (*Tg*). The gene is expressed in the thyroid gland and, as it has been recently demonstrated, forms so-called transcription loops, easily observable by light microscopy. Using this feature, we show that *Tg* is expressed at a high level from the moment a thyroid cell acquires its identity and both alleles remain highly active over the entire life of the cell, i.e., for months or years depending on the species. We demonstrate that this high upregulation is characteristic of thyroglobulin genes in all major vertebrate groups. We provide evidence that *Tg* is not influenced by the thyroid hormone status, does not oscillate round the clock and is expressed during both the exocrine and endocrine phases of thyrocyte activity. We conclude that the thyroglobulin gene represents a unique and valuable model to study the maintenance of a high transcriptional upregulation.

KEYWORDS

thyroglobulin gene, transcription loop, transcription, gene upregulation, thyroid hormones

Introduction

It has been recently shown that highly expressed genes expand from their harboring loci and form so called transcription loops (TLs) (Mirny and Solovei, 2021). The expansion of highly expressed genes is attributed to their intrinsic stiffness acquired through decoration of the gene axis with multiple jam-packed RNA polymerases II (RNAPIIs) with attached nascent RNA transcripts (Figure 1A). Therefore, not only the gene length but also the high gene expression level are essential for formation of a microscopically resolvable transcription

10.3389/fcell.2023.1265407

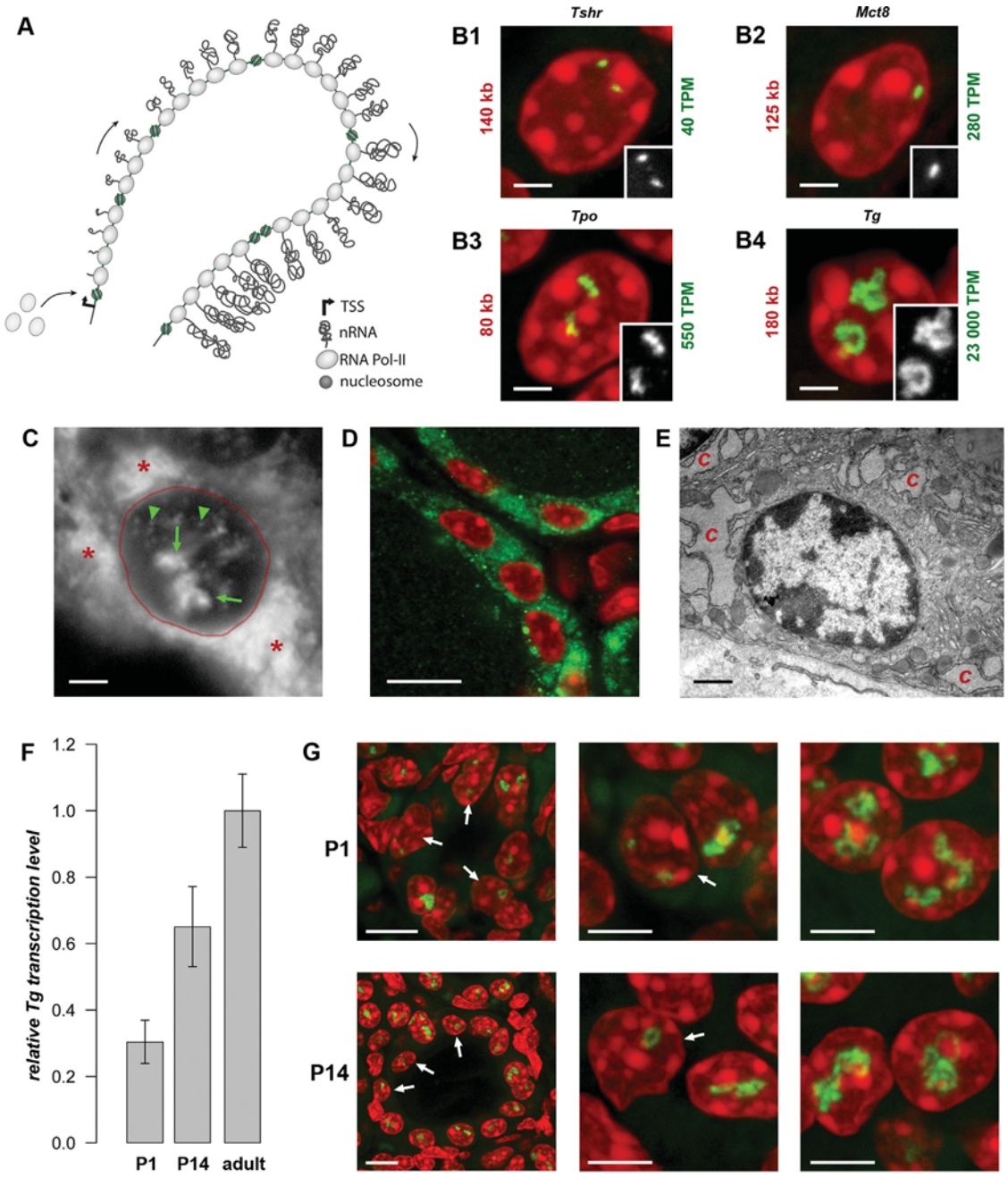


FIGURE 1
Transcription loops. **(A)** Schematics of a transcription loop (TL) formed by RNAPIIs moving along a gene and carrying nascent RNA transcripts. **(B)** Visualization of four thyroid-specific genes by RNA-FISH. In contrast to the lowly expressed genes *Tshr* and *Mct8* **(B1, B2)**, the other two genes, *Tpo* and *Tg* **(B3, B4)**, exhibit larger TLs with sizes correlated to their expression levels. Length and expression level of the genes are indicated on the left and right of the panels, respectively. For better comparison, TLs are shown as grey scale images in the insertions. Note, that in case of the *Mct8* gene located on X chromosome, there is only one signal, because the tissue originated from a male mouse. For the number of analysed nuclei, see Table 1. **(C)** RNA-FISH with an oligoprobe for *Tg* mRNA labels TLs (*green arrows*), single mRNAs in the nucleoplasm (*green arrowheads*) and in the cytoplasm (*red asterisk*). **(D,E)** Immunostaining of TG (*green*) highlights the cytoplasm of thyrocytes **(D)**, which is densely packed with remarkably hypertrophic cisterns **(C)** of endoplasmic reticulum **(E)**. **(F)** Histogram showing comparative *Tg* transcript levels in thyroids of young mice (P1 and P14) in comparison to adult mice. Since thyroid tissue includes 40% of non-thyrocyte cells, as well as chunks of practically inseparable parathyroid gland, the qPCR values were normalized to the transcription level of the thyroid specific transcription factor Pax8 to avoid erroneous results. Error bars are SDM, three biological replicates were analyzed per developmental stage The observed changes were not significant ($p > 0.05$) as determined by one-way ANOVA with Tukey's multiple comparisons *post hoc*. **(G)** Examples of thyrocytes from P1 (*top*) and P14 (*bottom*) thyroids. The *left column* shows follicles formed by thyrocytes with various degree of *Tg* TL development; the *right and middle columns* show thyrocytes at a higher magnification with fully or underdeveloped *Tg* TLs. Arrows point at nuclei with underdeveloped loops. Note that in some nuclei only one allele is active. For P1 and P14, image stacks of about 50 and 70 nuclei, respectively, were acquired. Images on B-D and G are projections of 3–5 µm confocal stacks; RNA-FISH and immunostaining signals are *green*; nuclei are counterstained with DAPI (*red*). Scale bars: B, C, E, 2 µm; D, 70 µm; G, follicle overviews on the left, 10 µm, zoomed in nuclei, 5 µm.

loop (Leidescher et al., 2022). Exactly this combination of gene properties is extremely rare, especially in mammalian cells. Consequently, TLs had not been described until the recent serendipitous discovery of TLs formed by the unusually upregulated thyroglobulin gene.

The thyroglobulin gene (*Tg*) is expressed exclusively in thyrocytes, which in vertebrates are arranged in follicles that, in turn, united into thyroid glands in most vertebrate groups. The thyroglobulin protein (TG) accumulates in the follicle cavity, representing the major component of the so-called colloid, and serves as a long-storage precursor for synthesis of the thyroid hormones (THs), 3,3′,5,5′-Tetraiodo-L-thyronine (T4) and 3,3′,5′-Triiodo-L-thyronine (T3) (van de Graaf et al., 2001). Importantly, *Tg* is highly upregulated with a transcription level of ca. 23,000 TPM (Transcripts Per Million), exceeding by many fold the expression of other thyrocyte-specific genes, as well as housekeeping genes, e.g., ubiquitinase (6-fold), actin (9-fold), myosin (22-fold) and multiple ribosomal genes (from 10- to 20-fold). Presumably, *Tg* is transcribed in long transcription bursts separated by short infrequent pauses (Figure 1A). In agreement with this, Hi-C analysis of thyroid tissue revealed that *Tg* TL formation perturbs local cohesin-mediated organization by dissolving the *Tg* sub-TAD and reorganizing the larger TAD (see ED Figure 7 in Leidescher et al., 2022).

The presence of two highly extended *Tg* TLs, corresponding to the two active gene alleles in more than 93% of nuclei (Leidescher et al., 2022), indicates that the gene is remarkably upregulated in almost every thyrocyte for a long time. Therefore, we were interested in investigating thyroglobulin gene activity more closely. In particular, we followed *Tg* activation in development and showed that the *Tg* TL is a hallmark of functional thyrocytes. We asked whether the thyroglobulin gene in other vertebrates is similarly upregulated and demonstrated that thyrocytes of all major vertebrate groups exhibit thyroglobulin TLs. We further investigated whether *Tg* is regulated by organismal or cellular physiological conditions and proved that *Tg* transcription and translation are not regulated by the thyroid hormone (TH) status and do not undergo circadian rhythmicity or intron retention, ruling out these possible mechanisms of temporal segregation of the exocrine and endocrine phases.

Material and methods

Tissue collection

All mouse studies were executed in accordance with the European Union (EU) directive 2010/63/EU on the protection of animals used for scientific purposes and in compliance with regulations by the respective local Animal Welfare Committees LMU; Committee on Animal Health and Care of the local governmental body of the state of Upper Bavaria; Germany; Animal Welfare Committee of the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen (LANUV; Recklinghausen, Germany). CD-1 mice were purchased from Charles River Laboratories, housed in individual cages with free access to food and water on a 12:12 light dark cycle at the Biocenter, Ludwig-Maximilians-University of Munich (LMU). Mice were sacrificed by cervical dislocation after IsoFlo (Isofluran, Abbott) narcosis. *Mct8*-KO mice (Trajkovic et al., 2007), *Trhr1*-KO mice (Rabeler et al., 2004); and *Thrb*-KO mice (Forrest et al., 1996), all on C57BL/6 background, were kept at 22°C in IVC cages in the central animal facility of the University Hospital Essen and had access to normal chow and water *ad libitum*. Adult mutant mice and control littermates were killed by CO_2 inhalation prior to organ collection.

Human thyroid tissue was sampled from a freshly operated struma according to the ethical consent (ethical approval No. 12-5133-BO to DF) of the Department of Endocrinology, University Hospital Essen, Germany. Thyroids from chicken (*Gallus gallus domesticus*) were collected from white legorn birds. Fertilized eggs of the M11 chicken line were kindly provided by Dr. S. Weigend (Federal Research Institute for Animal Health, Mariensee) and hatched at the Faculty for Veterinary Medicine, Munich. Birds were housed under conventional conditions in aviaries with groups of up to 10 birds and received food and water *ad libitum*. The animals were treated according to the standard protocol approved by The Committee on Animal Health and Care of the local governmental body of the state of Upper Bavaria, Germany. Thyroids from *Xenopus tropicalis* (*Silurana tropicalis*) were collected from four-year-old adult animals. Frogs were purchased at the Centre de Ressources Biologiques de Rennes (CRB) and raised in the PhyMa Laboratory in aquatic housing system (MPAquarien, Rockenhausen, Germany) at 24°C. Generation of transgenic zebrafish (*Danio rerio*) lines expressing fluorophores in thyrocytes, tg (tg:mCherry) and tg (tg:GFP) is described elsewhere (Opitz et al., 2012; Opitz et al., 2013).

In each case, freshly dissected tissues were washed with PBS and then fixed with 4% paraformaldehyde (Carl Roth) solution in PBS for 12–20 h.

Mouse and human organoids

Generation of functional mouse and human thyroid tissues *in vitro* is described in (Antonica et al., 2012; Romitti et al., 2022). In our experiments we used thyrocyte follicles formed *in vitro* in 3D matrigel culture and follicles formed after grafting differentiated *in vitro* thyrocytes into mouse kidney. All grafting experiments were performed in accordance with local Animal Ethics [Commission d'Ethique du Bien-Être Animal (CEBEA) Faculté de Médecine ULB, Project CMMI-2020-01].

Primary thyrocyte culture

For cultivating mouse primary thyrocytes *in vitro*, the protocol from Jeker et al., 1999 was adapted. Briefly, thyroid glands of two mice, were minced into pieces using micro scissors under binocular and transferred into 2 mL tube containing 200 U/mL collagenase and 1 U/mL dispase in DMEM/F12/GlutaMax medium. Incubation was performed in a shaking thermo-block at 37°C for 2 h and followed by mechanical disruption using a glass Pasteur pipetting. Then cells were centrifuged at 2,000 rpm (358 g) for 5 min and resuspended in DMEM supplemented with 10% FCS and 3% Pen/Strep. Single thyrocytes and follicles were seeded on coated coverslips (pre-incubated with 1 µg/mL polylysine) and

cultured for 30 min, 24 h or 72 h before fixation with 4% formaldehyde.

Cryosections

After fixation, thyroids were washed with PBS, cryoprotected in a series of sucrose, and embedded in Tissue-Tek O.C.T. compound freezing medium (Sakura). Blocks were stored at −80°C before cutting into 16–20 µm sections using a cryostat (Leica CM3050S). Cryosections were collected on Superfrost Plus slides (Thermo Scientific) and stored at −80°C before use.

Electron Microscopy

Mouse thyroid glands were fixed with 2% glutaraldehyde in 300 mOsm cacodylate buffer (75 mM cacodylate, 75 mM NaCl, 2 mM MgCl2) for 30 min, postfixed with 1% OsO4 in the same buffer for 1 h at room temperature. After washings in distilled water, samples were incubated in 1% aqueous solution of uranyl acetate (Serva) for 1 h at 4°C, dehydrated in ethanol series and acetone, and embedded in Epon Resin. Thin sections (50–70 nm) were prepared using Reichert Ultracut, stained with Reynolds lead citrate and examined with a transmission electron microscope (JEM 100 SX, JEOL) at 60 kV.

Gene expression analysis

Dissected thyroids were immediately placed in RNAlater solution and total RNA was isolated using the NucleoSpin RNA Kit (Macherey-Nagel) according to the manufacturer's instructions. RNA integrity was checked by separating RNA fragments on a 1% agarose gel. Only samples with a 28S:18S rRNA ratio of ~2:1 and without genomic contamination and RNA degradation were used for downstream applications. 1 µg of total RNA was reverse transcribed according to the manufacturer's instructions using the High capacity cDNA reverse transcription Kit with random primers (Applied Biosystems) or Maxima H Minus Reverse Transcriptase (Thermo Scientific) with gene specific primers. qPCR was performed in technical and biological triplicates in 10 µL reactions using LightCycler 480 SYBR Green Master Mix (Roche) or Luna Universal qPCR Master Mix (New England Biolabs) according to the manufacturer's instructions.

Nanopore sequencing

Freshly dissected thyroids were immediately placed in ice cold TriZol and homogenized using an ultra-turrax dispersing tool. Poly(A)+ transcripts were isolated using magnetic oligoT-beads (Lexogen). The sequencing library was created using the PCR-cDNA Sequencing Kit (PCB111.24, Oxford Nanopore) and sequenced on a PromethION P24 on a R9.4.1 flowcell. The sequencing data were basecalled using Guppy v6.4.6 and mapped to the mouse genome (mm10) using minimap2.

FISH probes

BAC clones encompassing the thyroglobulin genes and flanking regions of different vertebrate species, as well as *Tshr*, *Slc16a2* and *Tpo* mouse genes, were selected using the UCSC genome browser (see the list of BACs in Table 1) and purchased from BACPAC Resources (Oakland children's hospital) as agar stabs (https://bacpacresources.org/). BACs were purified via standard alkaline lysis or the NucleoBond Xtra Midi Kit (Macherey-Nagel), followed by amplification with the GenomiPhi Kit (GE Healthcare) according to the manufacturer's instructions. Amplified BAC DNA was labeled with fluorophores using homemade conjugated fluorophore-dUTPs by nick translation (Cremer et al., 2008). Labeled BAC DNA was ethanol precipitated with 10-fold excess of Cot-1 (1 mg/mL; Invitrogen, 18440-016) and 50-fold excess of salmon sperm DNA (5 µg/µL; Sigma), pellet was dried in a SpeedVak, and dissolved in hybridization mixture containing 50% formamide, 1xSSC and 10% of dextran sulphate. Oligoprobe for the *Tg* mRNA was generated using SABER-FISH protocol and described in detail previously (Leidescher et al., 2022).

FISH and immunostaining

FISH on cryosections was performed as previously described (Solovei, 2010). For DNA-FISH, sections were treated with 50 µg/mL RNaseA at 37°C for 1 h. For DNA-FISH or FISH detecting DNA and RNA simultaneously, denaturation of both probe and sample DNA was carried out on a hot block at 80°C for 3 min. For only RNA-FISH, RNasing and denaturation steps were omitted. Probes were loaded on sections under small glass chambers and sealed with rubber cement [for detail, see (Solovei, 2010)]. Hybridizations were carried out in a water bath at 37°C for 2 days. After hybridization, rubber cement and chambers were removed, slides with sections were washed with 2xSSC at 37°C, 3 × 30 min, and then with 0.1xSSC at 60°C 1 × 7 min. Hybridized SABER probes were detected by incubating with 1 µM fluorescently labeled detection oligonucleotides in PBS for 1 h at 37°C followed by washing with PBS for 10 min.

The primary antibody for cytoplasmic TG detection (1:50; rabbit anti-TG Abcam, ab 156008) and the secondary donkey anti-rabbit conjugated with Alexa 555 (1:250; Invitrogen, Cat# A31570) were diluted in blocking solution (PBS +2% BSA +0.1% Saponin +0.1% TritonX100) and applied under glass chambers covering sections. Incubations with primary and secondary antibodies were carried overnight at RT, in between and after incubations, sections were washed with PBS +0.05% TritonX100 warmed up to 37°C; 3 × 30 min. In all experiments, nuclei were counterstained with 2 µg/mL DAPI in PBS for 30 min and Vectashield (Vector) was used as an antifade mounting medium.

Microscopy

Confocal 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 120 nm. Axial chromatic shift correction, as well as building single grey-

TABLE 1 BACs used in the study and number of imaged thyrocyte nuclei.

	Latin name	Genomic region	BAC #	Number of imaged nuclei
Mouse	*Mus musculus*	*Tg*, gene body	RP24-229C15	> 3,000
		Tshr, gene body	RP23-152C21	32
		Slc16a2, gene body	RP23-114F20	29
		Tpo, gene body	RP24-172O17	35
Human	*Homo sapience*	*TG*, gene body	RP11-806J1	70
		TG, 5′ half	RP11-844M6	
		TG, 3′ half	RP11-111C8	
		TG, 5′ flank	RP11-599I7	
		TG, 3′ flank	RP11-739E11	
Chicken	*Gallus domesticus*	*tg*, gene body	CH261-128F8	38
		tg, 5′ flank	CH261-88K23	
		tg, 3′ flank	CH261-85H20	
Frog	*Xenopus tropicalis*	*tg*, gene body	CH216-364N2	59
		tg, gene body	CH216-26O19	
Zebrafish	*Danio rerio*	*tg*, gene body	CH211-184D12	30
		tg, 3′ 3/4 of the gene	CH73-56C22	

scale stacks, RGB-stacks, montages and maximum intensity projections was performed using ImageJ plugin StackGroom (Walter et al., 2006). The plugin is available upon request.

Results and discussion

Tg is an exceptionally highly expressed gene in thyrocytes

In agreement with its length (180 kb in mouse) and level of transcription (23,000 TPM), the *Tg* gene forms very prominent transcription loops (Figure 1B4). To examine other long thyroid-specific genes, we selected thyroid peroxidase (*Tpo*) with a length of 78 kb, the TH transporting monocarboxylate transporter MCT8 (*Slc16a2*) that is 125 kb long and the thyroid stimulating hormone receptor (*Tshr*) with a length of 139 kb. The RNA signals of *Slc16a2* and *Tshr*, which have comparable length and ca. 100-500 times lower expression levels than *Tg*, are small and only slightly elongated. On the contrary, the much shorter *Tpo* gene, which is about 80 kb long but expressed at ca. 550 TPM, noticeably expands and forms short but distinctive transcription loops (Figure 1B). This finding confirms that formation of microscopically resolvable transcription loops is dependent not only on the gene length, but also on the intensity of transcription (Leidescher et al., 2022).

Remarkably, every single thyrocyte in mouse thyroid manifests *Tg* transcription, and the majority of cells possess two TLs, with only 6.7% of thyrocytes exhibiting monoallelic expression (Leidescher et al., 2022). The high transcription of the *Tg* gene is consistent with high production of the thyroglobulin protein (TG). The *Tg* transcripts comprise 2.5% of the entire mRNA pool in mouse thyroid, similarly to 2.6% shown for human thyrocytes (Pauws et al., 2000). In accordance, the probe for *Tg* mRNA highlights not only *Tg* TLs and intranuclear mRNA, but also the whole thyrocyte cytoplasm (Figure 1C). In agreement with this, the cytoplasm is brightly stained with an antibody against TG (Figure 1D), and exhibits dilated cisternae of the Golgi apparatus, as well as a remarkable hypertrophy of the rough endoplasmic reticulum, which occupies most of the thyrocyte cytoplasm (Figure 1E). The above observations corroborate the statement that the *Tg* gene is abnormally upregulated and thyrocytes produce immense amounts of *Tg* mRNAs and TG protein.

Thyrocytes exhibit *Tg* TLs from the onset of their differentiation

In mouse, the onset of folliculogenesis starts at E15.5 and the thyroid gland forms before birth (De Felice and Di Lauro, 2011). We aimed at estimating the *Tg* gene expression in early postnatal development and performed qPCR on thyroid glands dissected at stages P1 and P14. We showed that *Tg* expression increases during development, reaching a 3.3-fold rise in the adult mice compared to P1 pups (Figure 1F). RNA-FISH revealed that although *Tg* TLs in most P1 and P14 thyrocytes are fully developed, ca. 20% and 12% of the cells, respectively, possessed small, undeveloped loops and often only one active allele (Figure 1G, arrows). Apparently, these cells represent freshly differentiated or still differentiating thyrocytes. Although mitotic cells in the follicle epithelium are relatively infrequent, in both developmental stages, we observed several thyrocytes entering or exiting the cell cycle. In particular, we noticed that *Tg* TLs are manifested when most of the nuclear

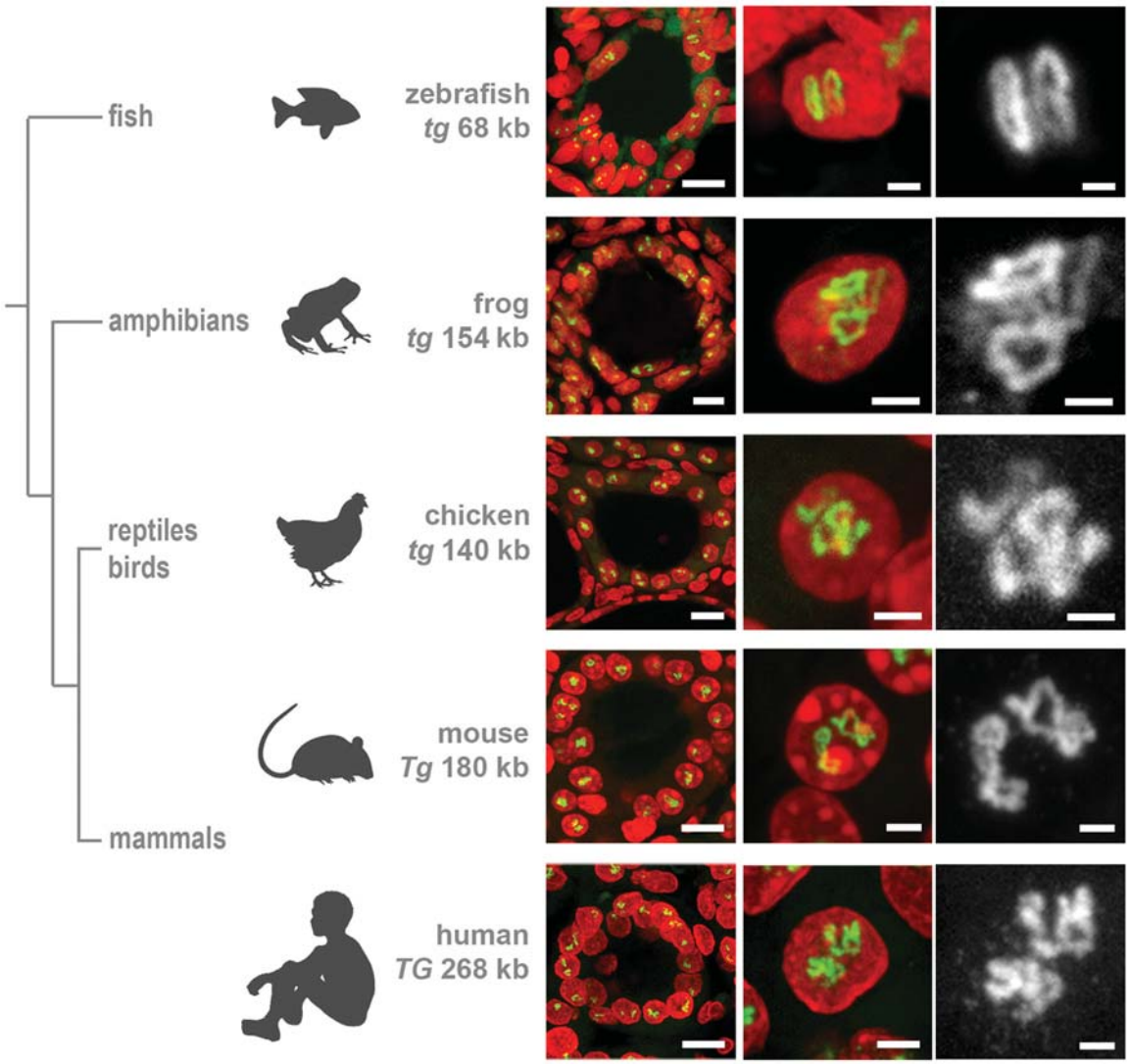


FIGURE 2
Thyroglobulin TLs are a peculiarity of thyrocytes from all major vertebrate groups. RNA-FISH with species-specific probes hybridizing to nascent RNA transcripts of thyroglobulin genes reveals TLs in all studied species. *The left column* shows thyroid follicles; *the mid column* shows representative thyrocyte nuclei at a higher magnification, *the right column* shows close-ups of *Tg* TLs of the middle column as greyscale images. Note that even the shortest among vertebrates, zebrafish *tg* with a length of only 68 kb, forms TLs resolvable by light microscopy. For the number of analyzed nuclei for each species, see Table 1. RNA-FISH signals (*green*), DAPI (*red*); images are projections of confocal 3–6 µm stacks. Scale bars: left column, 10 µm; mid column, 2 µm; right column, 1 µm.

chromatin is condensed: they appear in early G1, remain visible in thyrocyte nuclei until early-to-mid prophase and are withdrawn only in the very late prophase (Supplementary Figure S1). Thus, we conclude, that the *Tg* gene is expressed at a high level from the moment cells acquire thyrocyte identity.

The *Tg* gene is highly expressed and forms transcription loops in all vertebrates

Initially, the thyroglobulin TLs have been studied in mouse thyrocytes (Leidescher et al., 2022). However, THs are essential regulatory elements in the development and metabolism of all animals, and the mechanisms of TH generation are evolutionary conserved (Di Jeso and Arvan, 2016; Holzer et al., 2016). Therefore, we anticipated that in other mammals and other vertebrate classes, expression of thyroglobulin gene orthologues is also highly upregulated.

First, we demonstrated that human *TG* (length of 280 kb) exhibits TLs in thyrocytes of a freshly operated human thyroid (Figure 2). Next, we detected the thyroglobulin gene (*tg*) in the chicken *Gallus domesticus* (length of ca.140 kb) and the frog *Xenopus tropicals* (length of ca.154 kb), verifying that the gene forms TLs in these species as well (Figure 2). Finally, we aimed at visualization of the thyroglobulin gene in fish species, where thyrocytes are assembled into single follicles, which are not gathered

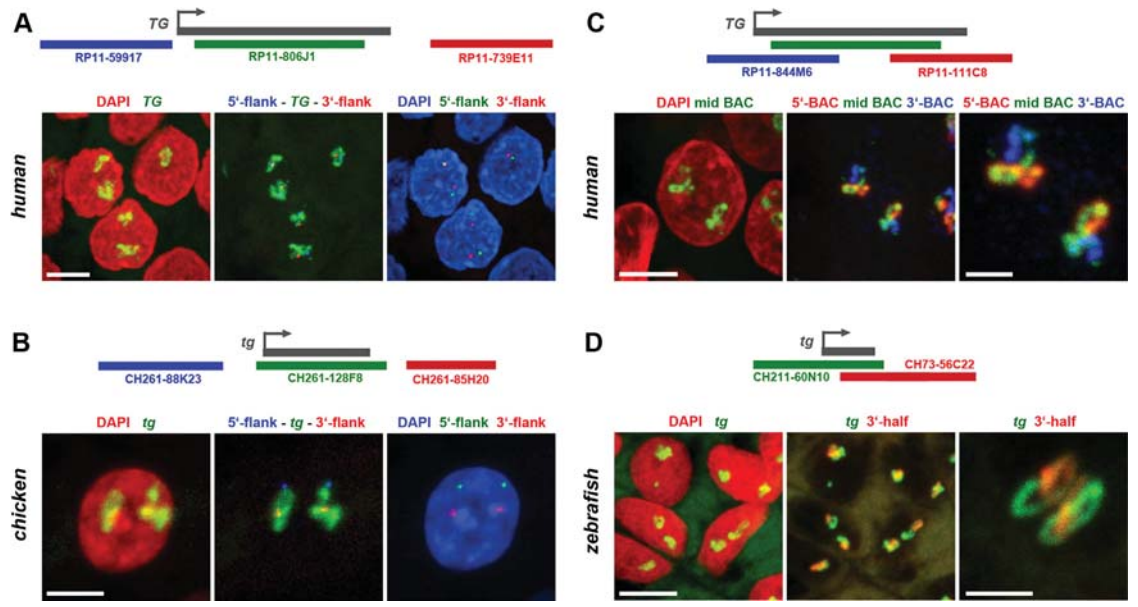


FIGURE 3
Thyroglobulin TLs in other vertebrates exhibit the key features of TLs. **(A,B)** As exemplified by simultaneous RNA- and DNA-FISH for human **(A)** and chicken **(B)** thyroglobulin, their TLs are open loops with visibly separated flanks. Images on the *left* show thyrocyte nuclei (*red*) with entire TLs (*green*). *Mid* images show the same TLs in *green* and the two flanks in *blue* (5′) and *red* (3′). Images on the *right*, for better demonstration of flank separation, show the same nuclei (blue) and the two flanks in *green* (5′) and *red* (3′). **(C,D)** As exemplified by RNA-FISH for human **(C)** and zebrafish **(D)** thyroglobulin genes, genomic probes highlighting introns, label TLs sequentially as a result of a co-transcriptional splicing. *Left* images show thyrocyte nuclei (*red*) with entire TLs (*green*). *Mid* images show the same TLs after hybridization with mid (*green*), 5′ (*blue*) and 3′ (*red*) BACs **(C)**. For zebrafish *tg*, only two BACs were used, one covering the whole gene (*green*) and one—its 3′ half (*red*) **(D)**. Images on the *right* show TLs at a higher magnification with the same colour code. Between 20 and 40 nuclei were acquired for each of the experiments. Schematics of genes and used BACs are shown above every panel. Images are projections of confocal stacks through 6 µm **(A,C)**, 3.5 µm **(B)** and 2.5 µm **(D)**. Scale bars: A-D, 5 µm; right panels on C and B, 2 µm.

into a gland, remain separated and are scattered along the esophagus. Therefore, to overcome the difficulties of finding follicles, we made use of the transgenic zebrafish *Danio rerio* lines generated previously, in which thyrocytes express fluorophores under the *tg* promoter, tg (*tg*:mCherry) or tg (*tg*:GFP) (Opitz et al., 2012; Opitz et al., 2013). Using cryosections from these transgenic lines, we demonstrated that, despite the relatively small length of the *D.rerio tg* (68 kb), it forms microscopically resolvable TLs strongly expanding throughout the nucleus (Figure 2).

Importantly, all the characteristic TL features, shown previously only for mouse TLs, are utterly manifested in other species as well. Firstly, we observed separation of TL flanks (Figures 3A, B), which has been previously described for other highly expressed long genes and is presumably explained by increased gene stiffness (Leidescher et al., 2022). Secondly, it has been demonstrated that probes highlighting introns over long gene stretches label TLs sequentially as a result of co-transcriptional splicing (Leidescher et al., 2022). This phenomenon is now also confirmed for human and fish thyroglobulin TLs by using sequential BAC probes hybridizing mostly to introns of nascent RNA transcripts decorating the loops (Figures 3C, D).

These data show a high upregulation of the thyroglobulin gene in the major vertebrate classes confirming the key role of TG in hormone production. This conclusion is reinforced by the recent work demonstrating a highly conserved structure of TG protein among vertebrates with all domains conserved from basal vertebrates to mammals (Di Jeso and Arvan, 2016; Holzer et al., 2016). Moreover, the authors of this paper conclude that since no orthologues of thyroglobulin gene have been found in invertebrates, TG protein and the mechanism for TH formation are vertebrate inventions.

Tg TLs are a robust mark of thyrocytes differentiated *in vitro*

Sabine Costagliola's research group described the generation of a functional thyroid tissue *in vitro* in 2012 (Antonica et al., 2012). The authors demonstrated that transient overexpression of the transcription factors Nkx2-1 and Pax8 in mouse embryonic stem cells, followed by 3D culturing, allows the generation of follicles similar in morphology and gene expression to the thyroid follicles *in vivo*. 10 years later, the same group using similar strategy succeeded in generating and growing human thyroid organoids (Romitti et al., 2022). To confirm the functionality of mouse and human follicles generated *in vitro*, cultured follicles of both species were grafted into kidneys of athyroid mice resulting in successful rescue of hormone production after 4–5 weeks.

Not surprisingly, RNA-FISH revealed *Tg* TLs in mouse and human follicles cultured *in vitro*, although the loop size was noticeably smaller in comparison to thyrocytes in glands (Supplementary Figure S2). To visualize mouse *Tg* TLs in the grafted thyrocytes, we used RNA- and DNA-FISH simultaneously—to identify injected male mouse cells in

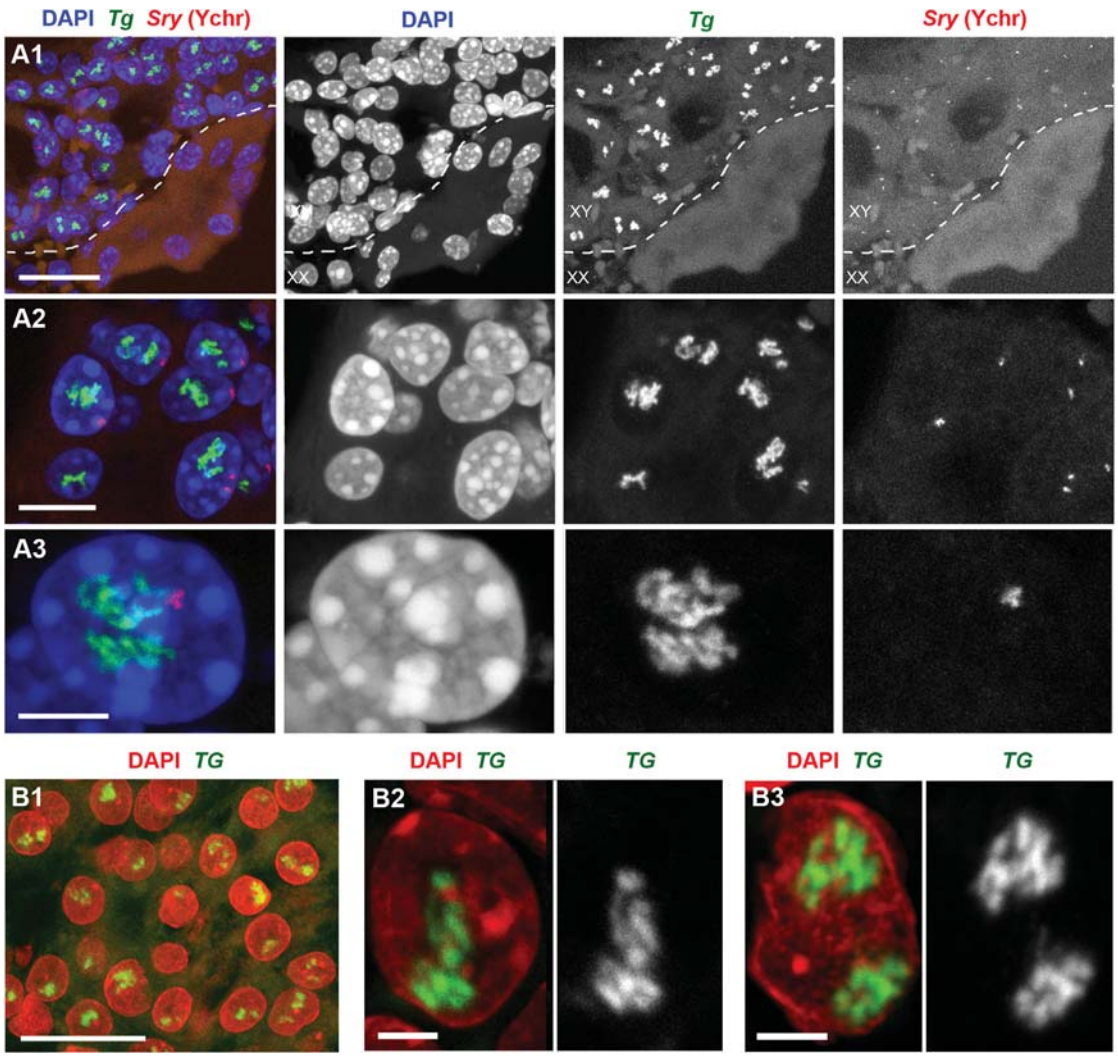


FIGURE 4
Mouse **(A)** and human **(B)** thyroid organoids grafted into mouse kidney. Cells of grafted follicles, generated from mouse male ESCs, can be distinguished from female mouse host cells by DNA-FISH with a Y chromosome-specific probe (*red*). The dotted line in **(A1)** separates host kidney cells from thyrocytes of follicles. Note the high extension of *Tg* and *TG* TLs in comparison to TLs formed by these genes in cultured follicles (compare to Supplementary Figure S2). 30 and 32 stacks were acquired from mouse and human grafted thyroid organoids, respectively. Thyroglobulin TLs are *green*; DAPI is *blue* in **(A)** and *red* in **(B)**. Images are projections of confocal 4 µm stacks **(A1,B1)** and 2.5 µm **(A2,A3,B2,B3)**. Scale bars: **(A1,B)** 25 µm, **(A2)** 10 µm, **(A3)** 5 µm, **(B2,B3)** 2.5 µm.

the female host tissue by a probe for the Y chromosome (DNA-FISH) and to detect *Tg* loops (RNA-FISH). Accumulations of small follicles were clearly distinguished from the host kidney tissue by Y chromosome DNA-FISH signals (Figure 4A), and predictably, only cells with this marker exhibited *Tg* TLs, which were much more extended compared to cultured follicles or even thyrocytes within the thyroid gland (Figures 4A2, 4A3).

Differentiation between host mouse cells and human cells grafted into mouse kidney is an easy task, because in difference to human, mouse nuclei possess multiple chromocenters formed by subcentromeric major satellite repeat (Vissel and Choo, 1989; Erdel et al., 2020) brightly stained with DAPI (Supplementary Figure S3A). In addition, transgenic differentiated thyrocytes express NKX2-GFP and thus could be distinguished from non-thyrocytes (Supplementary Figure S3B). Similarly to mouse, human grafted thyrocytes form follicles (Figure 4B1) and the *TG* genes become strongly expanded (Figures 4B2, 4B). What is more, expressed *TG* modifies the harboring chromosomal locus by pushing its flanks away from each other (Supplementary Figure S3C) in a manner described for other resolvable TLs (Leidescher et al., 2022). Interestingly, using probes for the gene body and flanks, we noticed that several thyrocytes had four instead of two *TG* TLs (Supplementary Figure S3C), indicating that either during culturing *in vitro* or after grafting, some human thyrocytes became tetraploid. The *TG* TLs exhibit another typical TL feature, the co-transcriptional splicing: the three consecutive overlapping genomic probes sequentially label the nascent RNAs decorating the gene because the introns are sequentially spliced out (Supplementary Figure S3D). Thus, the work with thyrocytes

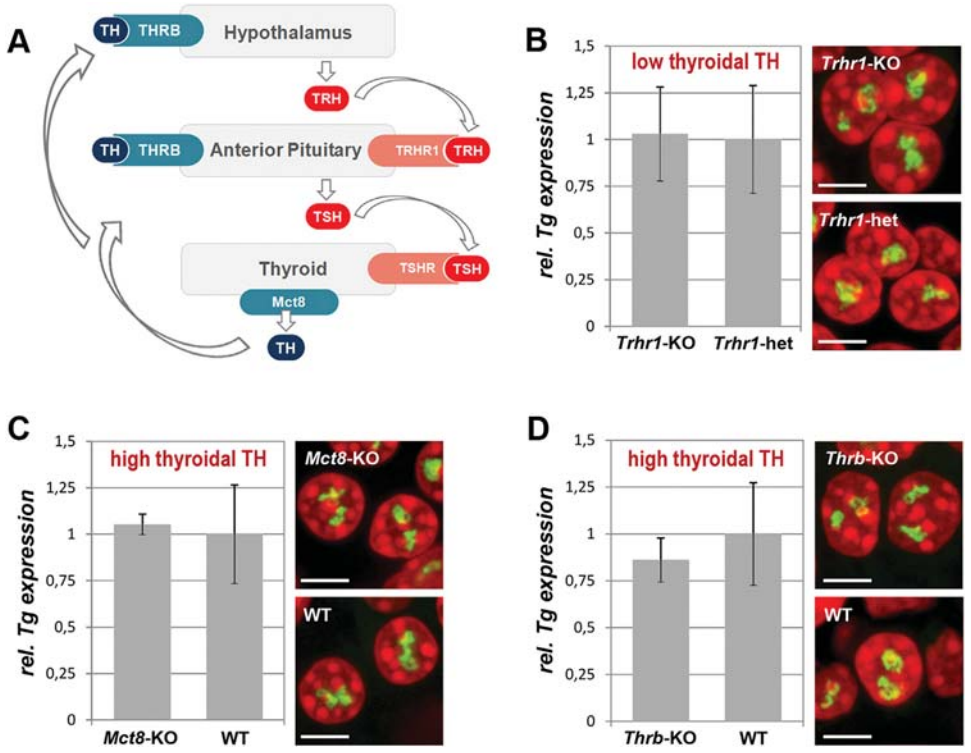


FIGURE 5
Thyroidal TH status does not influence *Tg* expression. **(A)** Simplified schematics of the hypothalamus-pituitary-thyroid axis. *TH*, thyroid hormones; *THRB*, TH receptor; *TRH*, hypothalamic thyrotropin-releasing hormone; *TRHR1*, thyrotropin-releasing hormone receptor; *TSH*, thyroid stimulating hormone; *TSHR*, thyroid stimulating hormone receptor. **(B–D)** Results of qPCR detecting levels of *Tg* expression (*left panels*) and RNA-FISH detecting *Tg* TLs (*right panels*) at a decreased **(B)** and increased **(C–D)** TH production in comparison to control mice. *Tg* transcript levels were normalized to the transcription level of the thyroid specific transcription factor Pax8 (see explanation in Figure 1F legend). Three biological replicates were analyzed per gene. All observed changes were not significant ($p > 0.05$) as determined by Wilcoxon rank sum test. For each of the conditions, two replicates were used for microscopy; between 20 and 30 confocal stacks with one or two nuclei were acquired. For more examples of nuclei, see Supplementary Figure S5. Bars in graphs are SDM. *Tg* TLs, *green*; DAPI, *red*; images are projections of 2–3 µm confocal stacks; scale bars: 5 µm.

differentiated from ESCs confirms that thyroglobulin TL formation is an invariable mark of differentiated thyrocytes.

Thyrocytes are functional only in follicles

The noticeably stronger extension of both *Tg* and *TG* TLs in grafts in comparison to thyrocytes *in vitro* indicates a higher upregulation of the gene within an organism. Culturing of single thyrocytes isolated from thyroid confirms the importance of the follicle structure for thyrocyte functional activity. Following the protocol by (Jeker et al., 1999) for thyroid disintegration and culturing (see Methods), we generated primary transient cultures of mouse thyrocytes, consisting mostly of single thyrocytes and remnants of follicles. Immediately after disintegration, cells were attached to coverslips, fixed and hybridized with a *Tg* genomic probe. RNA-FISH showed that thyrocytes exhibit *Tg* TLs for at least an hour after follicle disintegration (Supplementary Figures S4A, S4B). After 24 h of incubation, however, single thyrocytes became flatter and lost *Tg* TLs (Supplementary Figure S4C), although strongly reduced TLs were still present in some cells within the remaining flattened follicles. At this stage of the culture, no dividing cells were observed. After 72 h of incubation, all thyrocytes migrated out of follicles, became very flat, exhibited proliferative activity (with a mitotic index of 2.7%) and formed a monolayer, in which not a single cell exhibited *Tg* TLs (Supplementary Figure S4D). Apparently, thyrocytes lost their identity, possibly de-differentiated and entered the cell cycle as reported earlier for thyrocyte cultures established from other vertebrates (Kimura et al., 2001). These data indicate that results of various analyses conducted on cultured thyrocytes, as well as on other differentiated cells transferred to *in vitro* conditions, must be treated cautiously.

Tg expression is independent of thyroidal TH status

TH production is tightly regulated by the activity of the hypothalamus-pituitary-thyroid axis and controlled by negative feedback loops involving the TH receptor THRB (Ortiga-Carvalho et al., 2016). Hypothalamic thyrotropin-releasing hormone (TRH) activates its pituitary TRHR1 receptor and stimulates the thyroid by stimulating hormone (TSH) release, which in turn acts on TSHR of thyrocytes and stimulates the production and secretion of THs (Figure 5A). However, data on the regulation of the *Tg* gene activity are controversial. Earlier works

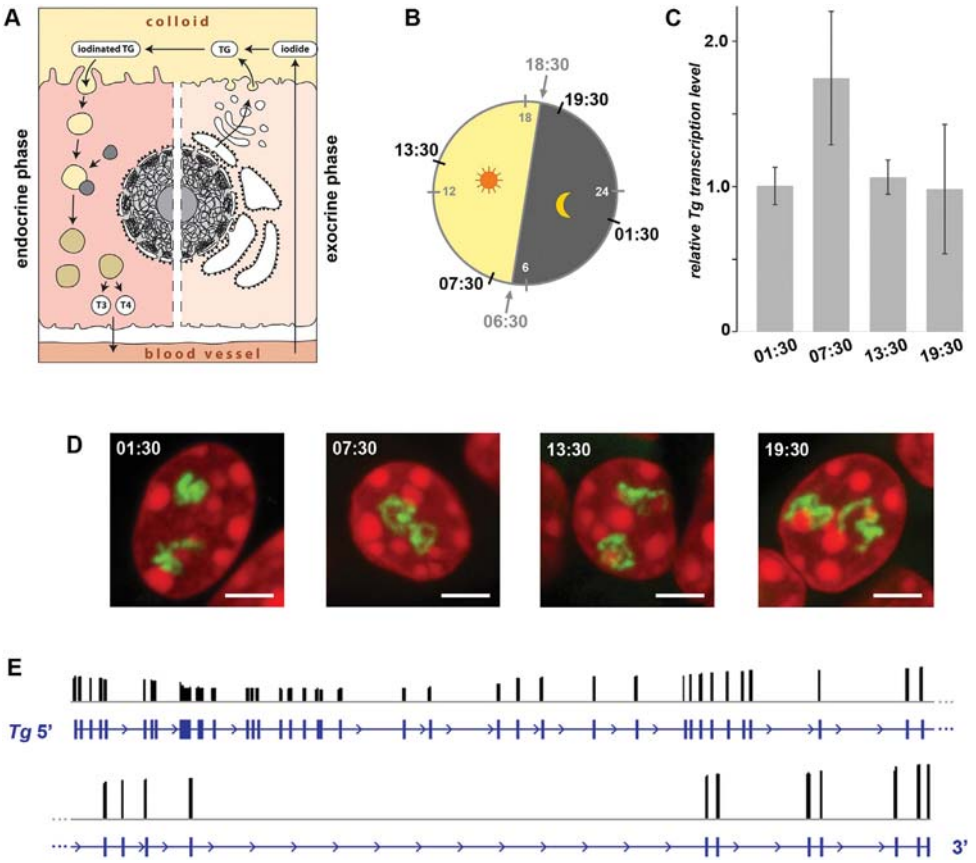


FIGURE 6
Expression of the *Tg* gene is not regulated by circadian activity and intron retention. **(A)** Schematics of a thyrocyte with two phases depicted - exocrine (*right*) and endocrine (*left*). **(B)** Diagram showing light: dark schedule (grey numbers) and time points (black numbers) of thyroid sampling for the circadian activity experiment. **(C)** Results of qPCR detecting levels of *Tg* expression indicate that there is no significant difference between the four time points. Bars are SDM. Expression levels were normalized to the 01:30 time point. Three biological replicates were analyzed per time point. The observed changes were not significant ($p > 0.05$) as determined by one-way ANOVA with Tukey's multiple comparisons *post hoc*. **(D)** Typical examples of thyrocyte nuclei after RNA-FISH detecting *Tg* TLs at four time points. *Tg* TLs, *green*; DAPI, *red*; images are projections of 3 µm confocal stacks; scale bars: 5 µm. For each of the three replicates at each time-point, between 40 and 50 confocal stacks were acquired. For more examples of nuclei, see Supplementary Figure S6. **(E)** Results of Nano-pore sequencing of poly(A) fraction for the *Tg* gene. RNA-seq read coverage (*black columns*) across the gene locus (*blue*). The read coverage is ranging between 0 and 3,146; only exon gene regions coincide with reads. For better presentation, the *Tg* gene is divided in two-halves that are shown one under the other.

showed that continuous presence of TSH is required to maintain TG production (Van Heuverswyn et al., 1984; Kim and Arvan, 1993). More recent work, however, showed that TSH deprivation or lack of functional TSHR due to early development does not affect the *Tg* expression but greatly reduces the expression of thyroperoxidase and the sodium/iodide symporter (Postiglione et al., 2002).

To test whether *Tg* expression is regulated by the thyroidal TH status, we sampled thyroids from mice with increased or decreased thyroidal TH production. For an increased thyroidal production we investigated *Mct8*-KO mice that exhibit a highly increased thyroidal TH content and reduced thyroidal T4 secretion (Trajkovic et al., 2007; Di Cosmo et al., 2010; Trajkovic-Arsic et al., 2010) (Figures 5A, C) as well as *Thrb*-KO mice that show highly elevated TSH and TH levels (Forrest et al., 1996) (Figures 5A, D). We also examined *Trhr1*-KO mice that display central hypothyroidism with decreased thyroidal and serum TH concentrations (Rabeler et al., 2004; Groba et al., 2013) (Figures 5A, B). For each condition, three thyroids were fixed for RNA-FISH and other three used for qPCR.

We anticipated to observe a drop of *Tg* transcription level and consequently a decrease in *Tg* TL size under "hyper" conditions, whereas "hypo" condition with decreased thyroidal TH concentration might show elevated *Tg* transcription level and thus increased *Tg* TL size. In all three conditions, however, *Tg* mRNA level did not differ from that of control samples and thyrocytes exhibited *Tg* TLs of the size and morphology similar to those in control animals (Figures 5B–D). Based on these results, we concluded that the intrathyroidal TH status does not influence *Tg* expression and that the gene is perpetually upregulated regardless of the activity of the hypothalamus-pituitary-thyroid axis.

Tg is upregulated during both the exocrine and endocrine activities of thyrocytes

The thyrocytes function as both exocrine and endocrine glands. On the apical side, a thyrocyte secretes proteins (e.g., TG, TPO) into

the follicle cavity filled with colloid, and on the basolateral site it releases TH into the circulation (Figure 6A). Surprisingly, the question whether these two phases of activity happen simultaneously or during separate time-windows remains open. One possibility to separate the phases would be a regulated oscillation in synthesis and excretion of TG, which is the major component of the follicle colloid, due to a potential *Tg* gene circadian rhythmicity. Indeed, human TSH exhibits a clear circadian rhythm with a peak between 2 and 4 a.m. (Russell et al., 2008). This fact suggests that the *Tg* gene might also be subject to circadian activity, separating in this way the two thyrocyte physiological phases.

To check the possible oscillation of the *Tg* gene, we assessed the abundance of *Tg* RNAs and the presence of *Tg* transcription loops in thyroid glands from healthy mice round-the-clock. The mice were entrained to a 12 h: 12 h light: dark schedule and sacrificed at four time points according to one of the standard schemes for circadian activity testing. The time points included 1 hour after the light switched on (07:30), 1 hour after the light switched off (19:30), and two time points in the middle (13:30 and 01:30) (Figure 6B). Thyroids of three animals were sampled for each time point, with one lobe of each thyroid used for qPCR to estimate *Tg* mRNA level and the other lobe used for microscopy.

In the first part of our analysis, using qPCR, we found that levels of Tg mRNA at the four time points were not significantly different (Figure 6C), indicating continuing high transcription and lack of acute waves of mRNA decay (Garneau et al., 2007). However, since the mRNA was extracted from entire cells (more accurately, from entire thyroid glands), our analysis included mixture of cytoplasmic and nucleoplasmic RNA fractions. Therefore, since Tg mRNA is super-abundant in the cytoplasm and apparently less so in nuclei, from such analysis we could not conclude that Tg transcription is not altered between the time points.

To clarify this question, in the second part of our analysis, we performed RNA-FISH highlighting nascent RNA transcripts and found that thyrocytes exhibit well developed Tg TLs at all four time points (Figure 6D and Supplementary Figure 6). Presence of TLs unlikely can be explained by "frozen" or "slowed down" transcription, because speed of RNAPII machinery is known to be pretty stable (3.8 kb/min($^{-1}$); Singh and Padgett, 2009), and therefore, microscopy control demonstrates that Tg transcription is indeed going on round the clock. According to the above data, we conclude that the Tg gene is not rhythmically expressed and remains highly upregulated throughout 24 h.

The presence of extended *Tg* TLs in thyrocytes round-the-clock still leaves an opportunity for the regulation of gene expression by accumulation of transcripts in the nucleoplasm followed by their concurrent export into the cytoplasm. Such mechanism of fine gene expression tuning is known as the intron retention phenomenon described for various cell types (Braunschweig et al., 2014). Intron retention allows an acute release of mRNA into the cytoplasm by synchronous excision of retained introns in response to a stimulus (Mauger et al., 2016). Thus, despite the perpetual transcription of *Tg*, intron retention might be a mechanism for *Tg* mRNA accumulation within nuclei during the endocrine phase and their release during the exocrine phase. To test this hypothesis, we isolated RNA from three mouse thyroid glands and performed Nanopore sequencing. As shown in Figure 6E, we have not detected intron retention in the poly(A) fraction of RNA: all sequenced reads mapped exclusively to the gene exons, which indicates that introns of all *Tg* mRNAs are excised. Therefore, we can rule out intron retention as a mechanism for restricting the TG production window.

Conclusion

Collectively, our data suggest that both thyroglobulin alleles are perpetually upregulated in thyrocytes of the major vertebrate groups with a seemingly everlasting expression during the entire cell life. On one hand, it underscores the physiological importance of TH for proper organismal development and function. On the other hand, in view of the large volumes of TG stored within follicles, such high and enduring expression of the *Tg* gene, at any condition, at any age and round-the-clock seems counterintuitive and remains enigmatic.

Conceivably, it might reflect an inefficient way of hormone production evolved during vertebrate evolution. Indeed, only a small proportion of the 66 tyrosine residues of the thyroglobulin molecule becomes iodinated and only three or four TH molecules result from cleavage of one thyroglobulin molecule during thyrocyte endocrine activity (van de Graaf et al., 2001; Di Jeso and Arvan, 2016). Therefore, such wasteful thyroglobulin production might be the way to correct this unintelligible nature design. Another not mutually exclusive explanation of the phenomenon is that massive TG production is needed for storage of the rare trace element iodine (Crockford, 2009). One can speculate that binding to a large protein is a safe way of building an iodine reservoir within an organism.

Taking in account the very low turnover of thyrocytes (Dumont et al., 1992), we deduce that the thyroglobulin gene is perpetually active, e.g., for months in mouse and for years in human (Coclet et al., 1989). In this respect, the phenomenon of the thyroglobulin gene represents an attractive model to study transcription regulation, in particular, molecular mechanisms of high upregulation maintenance, chromatin dynamics, and kinetics of splicing.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE233457.

Ethics statement

The studies involving humans were approved by the Ethical approval No. 12-5133-BO to DF of the Department of Endocrinology, University Hospital Essen, Germany. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. The animal study was approved by the LMU; Committee on Animal Health and Care of the local governmental body of the state of Upper Bavaria,

Germany; Animal Welfare Committee of the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

SU: Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft. SL: Writing–review and editing, Formal Analysis, Investigation, Methodology, Visualization. YF: Investigation, Methodology, Validation, Visualization, Writing–review and editing. KT: Writing–review and editing, Formal Analysis, Investigation, Methodology, Validation, Visualization. J-BF: Methodology, Resources, Writing–review and editing. BK: Methodology, Resources, Writing–review and editing. FW: Writing–review and editing, Methodology, Resources. BM: Writing–review and editing, Methodology, Resources. DF: Writing–review and editing, Methodology, Resources. MR: Writing–review and editing, Methodology, Resources. SK: Writing–review and editing, Data curation, Investigation, Methodology, Software. HB: Writing–review and editing, Data curation, Investigation, Methodology, Software. HL: Data curation, Funding acquisition, Writing–original draft, Writing–review and editing. SC: Writing–original draft, Conceptualization, Data curation, Resources. HH: Conceptualization, Data curation, Investigation, Methodology, Resources, Validation, Writing–original draft, Writing–review and editing. IS: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work has been supported by the Deutsche Forschungsgemeinschaft grants (SP2202/SO1054/2, project # 422388934 to IS, SPP 2202/LE721/17-1, project # 422857584 to HL, and SFB1064, project # 213249687 to HL and IS; SPP 1629/HE3418/7-2 to HH); The Belgian National Fund for Scientific Research (FNRS) (PDR T.0140.14; PDR T.0230.18, CDR J.0068.22), the Fonds d'Encouragement à la Recherche de l'Université Libre de Bruxelles (FER-ULB) and the European Union's Horizon 2020 research and innovation program under grant agreement No. 825745 (to SC); The French National Research center (CNRS) and the Muséum National of Natural History (to J-BF.)

Acknowledgments

We are grateful to Maria Carmo-Fonseca (Faculdade de Medicina da Universidade de Lisboa) for fruitful discussions and David Hicks (University of Strasbourg) for advices about the circadian rhythm experiment. We thank Conny Niemann and Andreas Klingl (Biozentrum, LMU, Munich) for technical assistance with electron microscopy.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2023.1265407/full#supplementary-material

References

Antonica, F., Kasprzyk, D. F., Opitz, R., Iacovino, M., Liao, X. H., Dumitrescu, A. M., et al. (2012). Generation of functional thyroid from embryonic stem cells. *Nature* 491, 66–71. doi:10.1038/nature11525

Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., et al. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 24, 1774–1786. doi:10.1101/gr.177790.114

Coclet, J., Foureau, F., Ketelbant, P., Galand, P., and Dumont, J. E. (1989). Cell population kinetics in dog and human adult thyroid. *Clin. Endocrinol. (Oxf)* 31, 655–665. doi:10.1111/j.1365-2265.1989.tb01290.x

Cremer, M., Grasser, F., Lanctot, C., Muller, S., Neusser, M., Zinner, R., et al. (2008). Multicolor 3D fluorescence *in situ* hybridization for imaging interphase chromosomes. *Methods Mol. Biol.* 463, 205–239. doi:10.1007/978-1-59745-406-3_15

Crockford, S. J. (2009). Evolutionary roots of iodine and thyroid hormones in cell-cell signaling. *Integr. Comp. Biol.* 49, 155–166. doi:10.1093/icb/icp053

De Felice, M., and Di Lauro, R. (2011). Minireview: intrinsic and extrinsic factors in thyroid gland development: an update. *Endocrinology* 152, 2948–2956. doi:10.1210/en.2011-0204

Di Cosmo, C., Liao, X. H., Dumitrescu, A. M., Philp, N. J., Weiss, R. E., and Refetoff, S. (2010). Mice deficient in MCT8 reveal a mechanism regulating thyroid hormone secretion. *J. Clin. Invest* 120, 3377–3388. doi:10.1172/JCI42113

Di Jeso, B., and Arvan, P. (2016). Thyroglobulin from molecular and cellular Biology to clinical Endocrinology. *Endocr. Rev.* 37, 2–36. doi:10.1210/er.2015-1090

Dumont, J. E., Lamy, F., Roger, P., and Maenhaut, C. (1992). Physiological and pathological regulation of thyroid cell proliferation and differentiation by thyrotropin and other factors. *Physiol. Rev.* 72, 667–697. doi:10.1152/physrev.1992.72.3.667

Erdel, F., Rademacher, A., Vlijm, R., Tunnermann, J., Frank, L., Weinmann, R., et al. (2020). Mouse heterochromatin adopts digital compaction states without showing hallmarks of HP1-driven liquid-liquid phase separation. *Mol. Cell* 78, 236–249. doi:10.1016/j.molcel.2020.02.005

Forrest, D., Hanebuth, E., Smeyne, R. J., Everds, N., Stewart, C. L., Wehner, J. M., et al. (1996). Recessive resistance to thyroid hormone in mice lacking thyroid hormone receptor beta: evidence for tissue-specific modulation of receptor function. *EMBO J.* 15, 3006–3015. doi:10.1002/j.1460-2075.1996.tb00664.x

Garneau, N. L., Wilusz, J., and Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* 8, 113–126. doi:10.1038/nrm2104

Groba, C., Mayerl, S., Van Mullem, A. A., Visser, T. J., Darras, V. M., Habenicht, A. J., et al. (2013). Hypothyroidism compromises hypothalamic leptin signaling in mice. *Mol. Endocrinol.* 27, 586–597. doi:10.1210/me.2012-1311

Holzer, G., Morishita, Y., Fini, J. B., Lorin, T., Gillet, B., Hughes, S., et al. (2016). Thyroglobulin represents a novel molecular architecture of vertebrates. *J. Biol. Chem.* 291, 16553–16566. doi:10.1074/jbc.M116.719047

Jeker, L. T., Hejazi, M., Burek, C. L., Rose, N. R., and Caturegli, P. (1999). Mouse thyroid primary culture. *Biochem. Biophys. Res. Commun.* 257, 511–515. doi:10.1006/bbrc.1999.0468

Kim, P. S., and Arvan, P. (1993). Hormonal regulation of thyroglobulin export from the endoplasmic reticulum of cultured thyrocytes. *J. Biol. Chem.* 268, 4873–4879. doi:10.1016/s0021-9258(18)53477-3

Kimura, T., Van Keymeulen, A., Golstein, J., Fusco, A., Dumont, J. E., and Roger, P. P. (2001). Regulation of thyroid cell proliferation by TSH and other factors: A critical evaluation of *in vitro* models. *Endocr. Rev.* 22, 631–656. doi:10.1210/edrv.22.5.0444

Leidescher, S., Ribisel, J., Ullrich, S., Feodorova, Y., Hildebrand, E., Galitsyna, A., et al. (2022). Spatial organization of transcribed eukaryotic genes. *Nat. Cell Biol.* 24, 327–339. doi:10.1038/s41556-022-00847-6

Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted intron retention and excision for rapid gene regulation in response to neuronal activity. *Neuron* 92, 1266–1278. doi:10.1016/j.neuron.2016.11.032

Mirny, L. A., and Solovei, I. (2021). Keeping chromatin in the loop(s). *Nat. Rev. Mol. Cell Biol.* 22, 439–440. doi:10.1038/s41580-021-00337-x

Opitz, R., Antonica, F., and Costagliola, S. (2013). New model systems to illuminate thyroid organogenesis. Part I: an update on the zebrafish toolbox. *Eur. Thyroid. J.* 2, 229–242. doi:10.1159/000357079

Opitz, R., Maquet, E., Huisken, J., Antonica, F., Trubiroha, A., Pottier, G., et al. (2012). Transgenic zebrafish illuminate the dynamics of thyroid morphogenesis and its relationship to cardiovascular development. *Dev. Biol.* 372, 203–216. doi:10.1016/j.ydbio.2012.09.011

Ortiga-Carvalho, T. M., Chiamolera, M. I., Pazos-Moura, C. C., and Wondisford, F. E. (2016). Hypothalamus-pituitary-thyroid Axis. *Compr. Physiol.* 6, 1387–1428. doi:10.1002/cphy.c150027

Pauws, E., Moreno, J. C., Tijssen, M., Baas, F., De Vijlder, J. J., and Ris-Stalpers, C. (2000). Serial analysis of gene expression as a tool to assess the human thyroid expression profile and to identify novel thyroidal genes. *J. Clin. Endocrinol. Metab.* 85, 1923–1927. doi:10.1210/jcem.85.5.6532

Postiglione, M. P., Parlato, R., Rodriguez-Mallon, A., Rosica, A., Mithbaokar, P., Maresca, M., et al. (2002). Role of the thyroid-stimulating hormone receptor signaling in development and differentiation of the thyroid gland. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15462–15467. doi:10.1073/pnas.242328999

Rabeler, R., Mittag, J., Geffers, L., Ruther, U., Leitges, M., Parlow, A. F., et al. (2004). Generation of thyrotropin-releasing hormone receptor 1-deficient mice as an animal model of central hypothyroidism. *Mol. Endocrinol.* 18, 1450–1460. doi:10.1210/me.2004-0017

Romitti, M., Tourneur, A., De Faria Da Fonseca, B., Doumont, G., Gillotay, P., Liao, X. H., et al. (2022). Transplantable human thyroid organoids generated from embryonic stem cells to rescue hypothyroidism. *Nat. Commun.* 13, 7057. doi:10.1038/s41467-022-34776-7

Russell, W., Harrison, R. F., Smith, N., Darzy, K., Shalet, S., Weetman, A. P., et al. (2008). Free triiodothyronine has a distinct circadian rhythm that is delayed but parallels thyrotropin levels. *J. Clin. Endocrinol. Metab.* 93, 2300–2306. doi:10.1210/jc.2007-2674

Singh, J., and Padgett, R. A. (2009). Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16 (11), 1128–1133. doi:10.1038/nsmb.1666

Solovei, I. (2010). Fluorescence *in situ* hybridization (FISH) on tissue cryosections. *Methods Mol. Biol.* 659, 71–82. doi:10.1007/978-1-60761-789-1_5

Trajkovic, M., Visser, T. J., Mittag, J., Horn, S., Lukas, J., Darras, V. M., et al. (2007). Abnormal thyroid hormone metabolism in mice lacking the monocarboxylate transporter 8. *J. Clin. Invest* 117, 627–635. doi:10.1172/JCI28253

Trajkovic-Arsic, M., Muller, J., Darras, V. M., Groba, C., Lee, S., Weih, D., et al. (2010). Impact of monocarboxylate transporter-8 deficiency on the hypothalamus-pituitary-thyroid axis in mice. *Endocrinology* 151, 5053–5062. doi:10.1210/en.2010-0593

Van De Graaf, S. A., Ris-Stalpers, C., Pauws, E., Mendive, F. M., Targovnik, H. M., and De Vijlder, J. J. (2001). Up to date with human thyroglobulin. *J. Endocrinol.* 170, 307–321. doi:10.1677/joe.0.1700307

Van Heuverswyn, B., Streydio, C., Brocas, H., Refetoff, S., Dumont, J., and Vassart, G. (1984). Thyrotropin controls transcription of the thyroglobulin gene. *Proc. Natl. Acad. Sci. U. S. A.* 81, 5941–5945. doi:10.1073/pnas.81.19.5941

Vissel, B., and Choo, K. H. (1989). Mouse major (gamma) satellite DNA is highly conserved and organized into extremely long tandem arrays: implications for recombination between nonhomologous chromosomes. *Genomics* 5, 407–414. doi:10.1016/0888-7543(89)90003-7

Walter, J., Joffe, B., Bolzer, A., Albiez, H., Benedetti, P. A., Muller, S., et al. (2006). Towards many colors in FISH on 3D-preserved interphase nuclei. *Cytogenet Genome Res.* 114, 367–378. doi:10.1159/000094227

frontiers

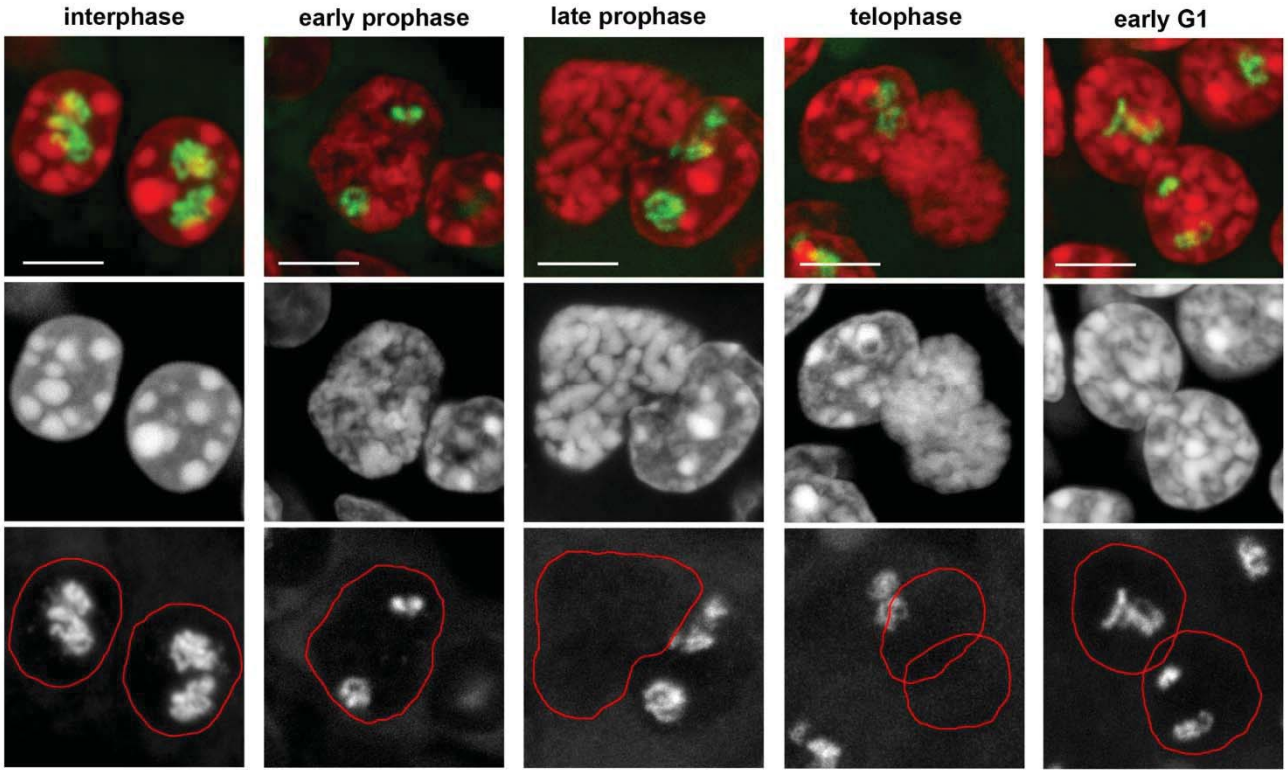
Supplementary Material

<u>Supplementary Figure 1</u>. The *Tg* gene in cycling thyrocytes.

TLs remain detectable by RNA-FISH until early-to-mid prophase, disappearing only during late prophase. The gene remains silent during metaphase, anaphase and early telophase, and becomes active in early G1. Note the *Tg* TLs are present when most chromatin is already (prophase) or still (G1) condensed.

Top panel shows overlay of *Tg* signals (*green*) and DAPI (*red*); *mid panel* shows grey scale images of DAPI counterstain; *bottom panel* shows FISH signals with nuclear borders outlined with a *red line*.

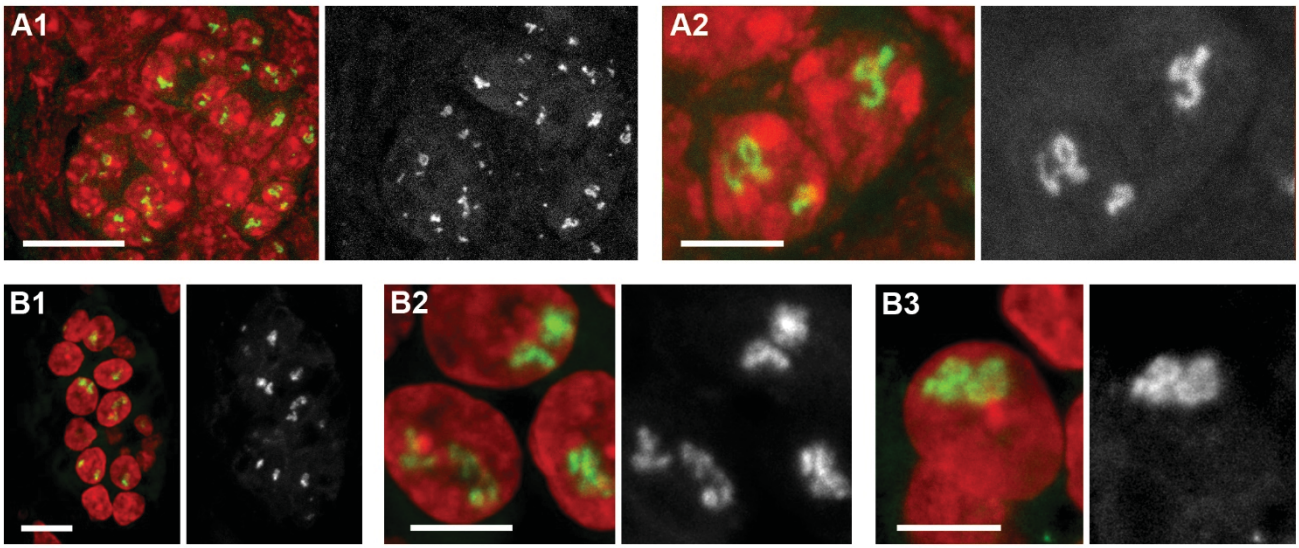
Images are projections of *3-6 µm* confocal stacks. Scale bars: *5 µm*.



<u>Supplementary Figure 2</u>. The *Tg* TLs in mouse (A) and human (B) organoids.

RNA-FISH detecting *Tg* TLs (*green*) in nuclei (*red*) of thyroid organoids. Follicles formed in culture with matrigel (A1, B1) and thyrocytes within the follicles at a higher magnification (A2, B2, B3). For clarity, every RGB panel is supplemented by grey scale images of TLs.

Images are projections of *4.5 µm* confocal stacks. Scale bars: A1,B1, *10 µm*, A2, B2, B3, *5* µm.



<u>Supplementary Figure 3</u> (appended to FIGURE 4). *TG* genes expressed in grafted follicles manifest typical TL features.

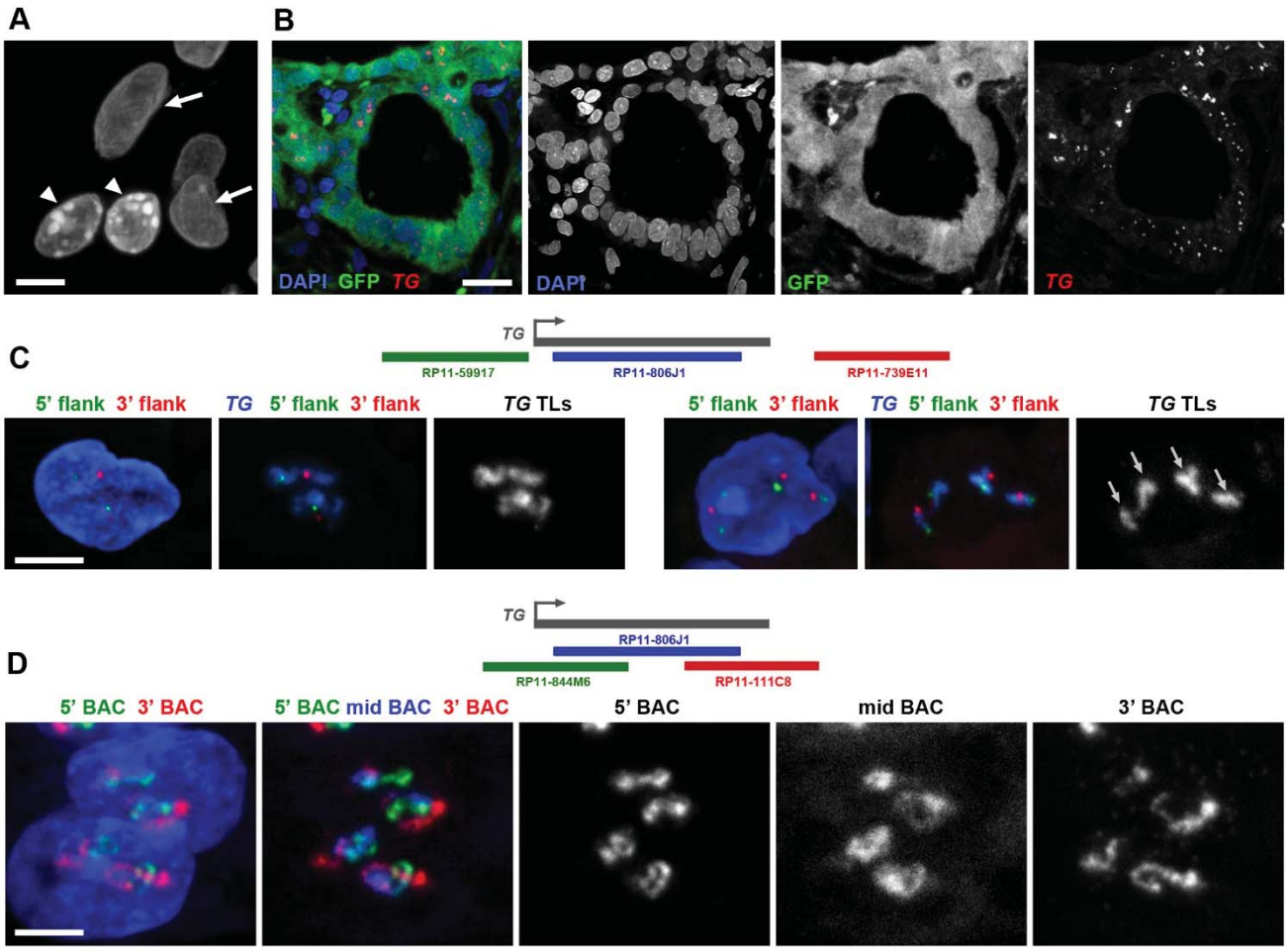
(A) Human thyrocytes (*arrows*) grafted into mouse kidney can be unambiguously distinguished from mouse cells (*arrowheads*), because they possess brightly stained with DAPI chromocenters.

(B) In addition, grafted human thyrocytes can be distinguished by the presence of GFP fluorescence (for detail, see Rommiti et al 2022).

(C) Flanks of *TG* TLs are visibly separated. Left panels show grafted thyrocyte nuclei and 5' (*green*) and 3' (*red*) flanks; mid panels show flanks and TLs in *blue*; right panels show only TLs as grey scale images. Note that the nucleus on the right is from a tetraploid human thyrocyte with four *TG* TLs marked with arrows.

(D) *TG* TLs are decorated by nRNA transcripts undergoing co-transcriptional splicing. Three BAC probes highlighting introns sequentially label *TG* loops. The left panel shows the nucleus in *blue*, the 5' BAC in *green* and the 3' BAC in *red*; the next RGB panel shows all three BACs including the middle one (*blue*). For clarity, the other three panels show grey scale images of each BAC signal.

Schematics of genes and used BACs are shown above C and D panels. Nuclei are counterstained with DAPI (*blue*). Images are projections of *2-3 μm* confocal stacks. Scale bars: B, *25 μm*, A,C,D, *5 μm*.

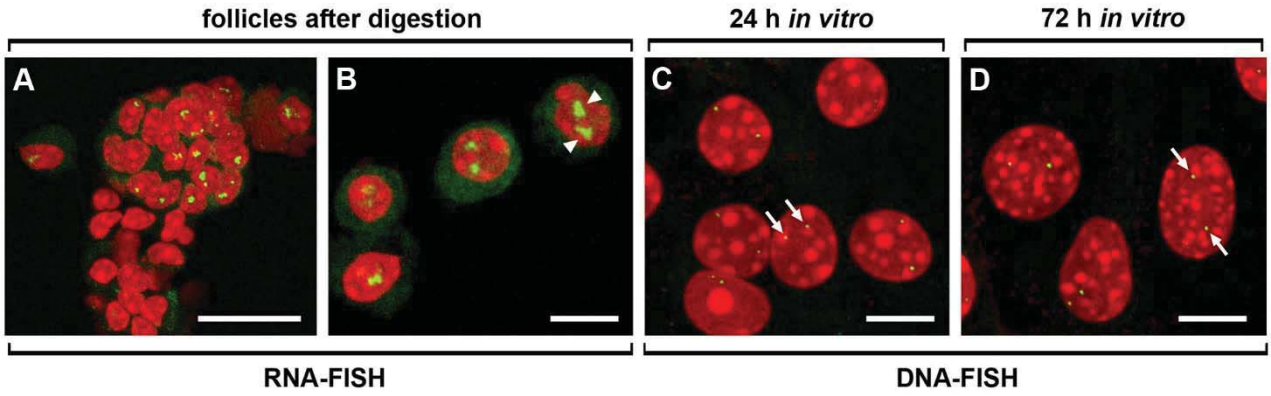


<u>Supplementary Figure 4</u>. Withdrawal of *Tg* TLs in primary thyrocyte culture.

(A, B) Freshly isolated follicle (A) and single thyrocytes (B) after RNA-FISH highlighting *Tg* TLs (*arrowheads*).

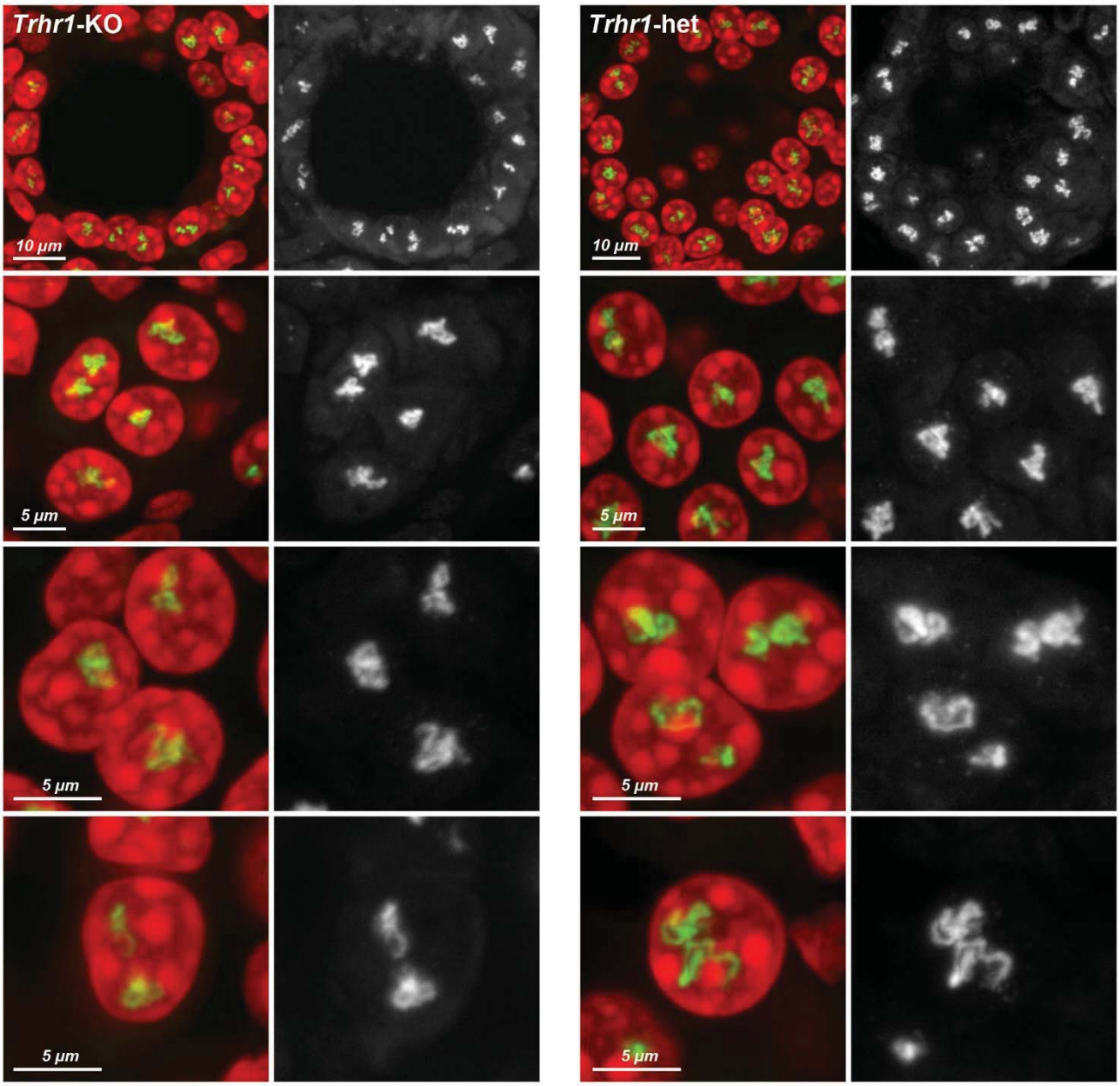
(C, D) After culturing for 1 (C) and 3 (D) days, mostly cubical thyrocytes flatten, enter the cell cycle and withdraw *Tg* TLs. Since the *Tg* genes become silenced and condensed, they can be detected only by DNA-FISH producing dot-like signals (*arrows*). Note the difference in the nuclear morphology of freshly isolated (B) and cultured for 3 days (D) cells.

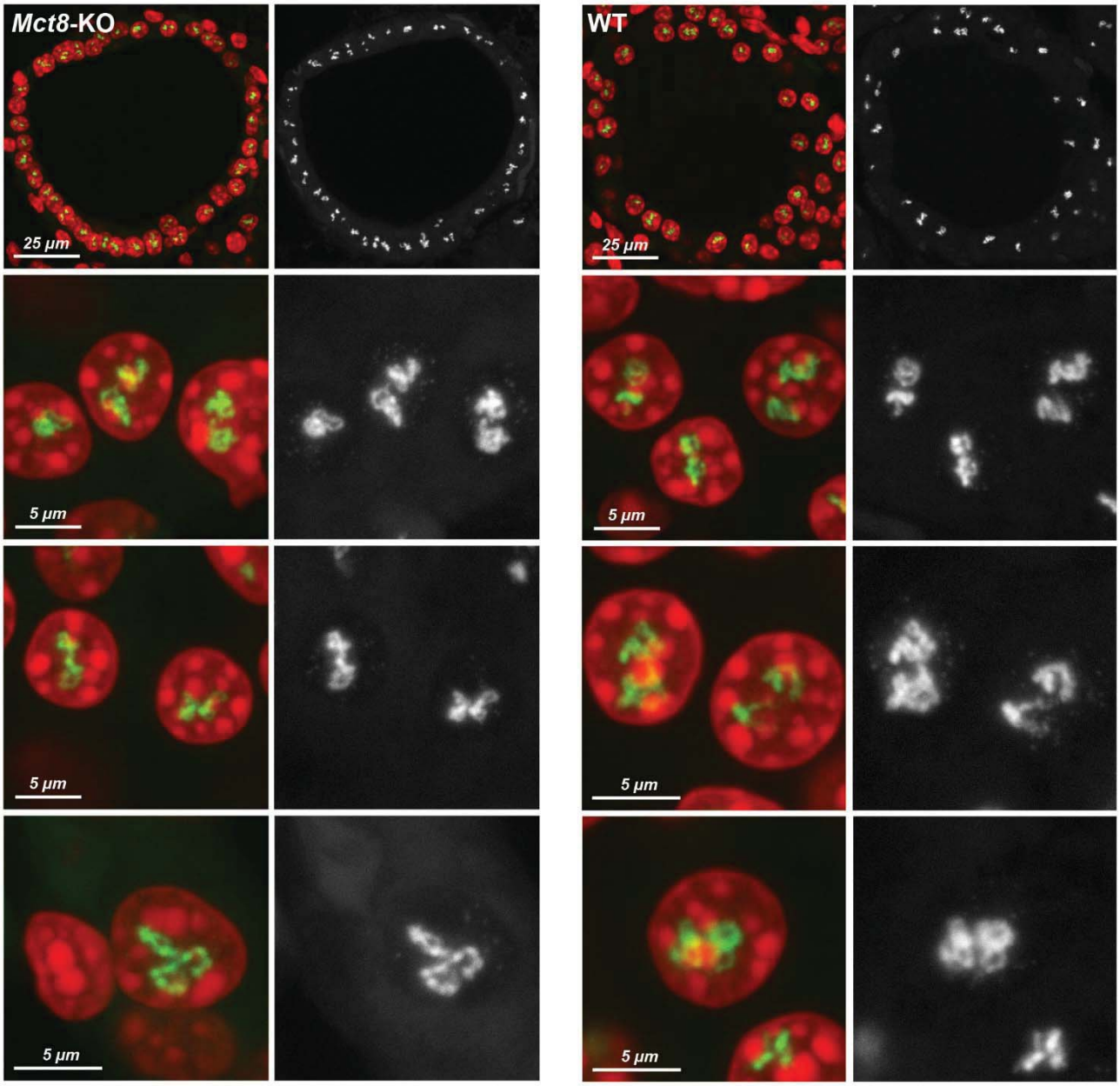
DAPI, *red*; RNA and DNA signals, *green*; images are projections of *3-5 µm* confocal stacks; scale bars, *10 µm*

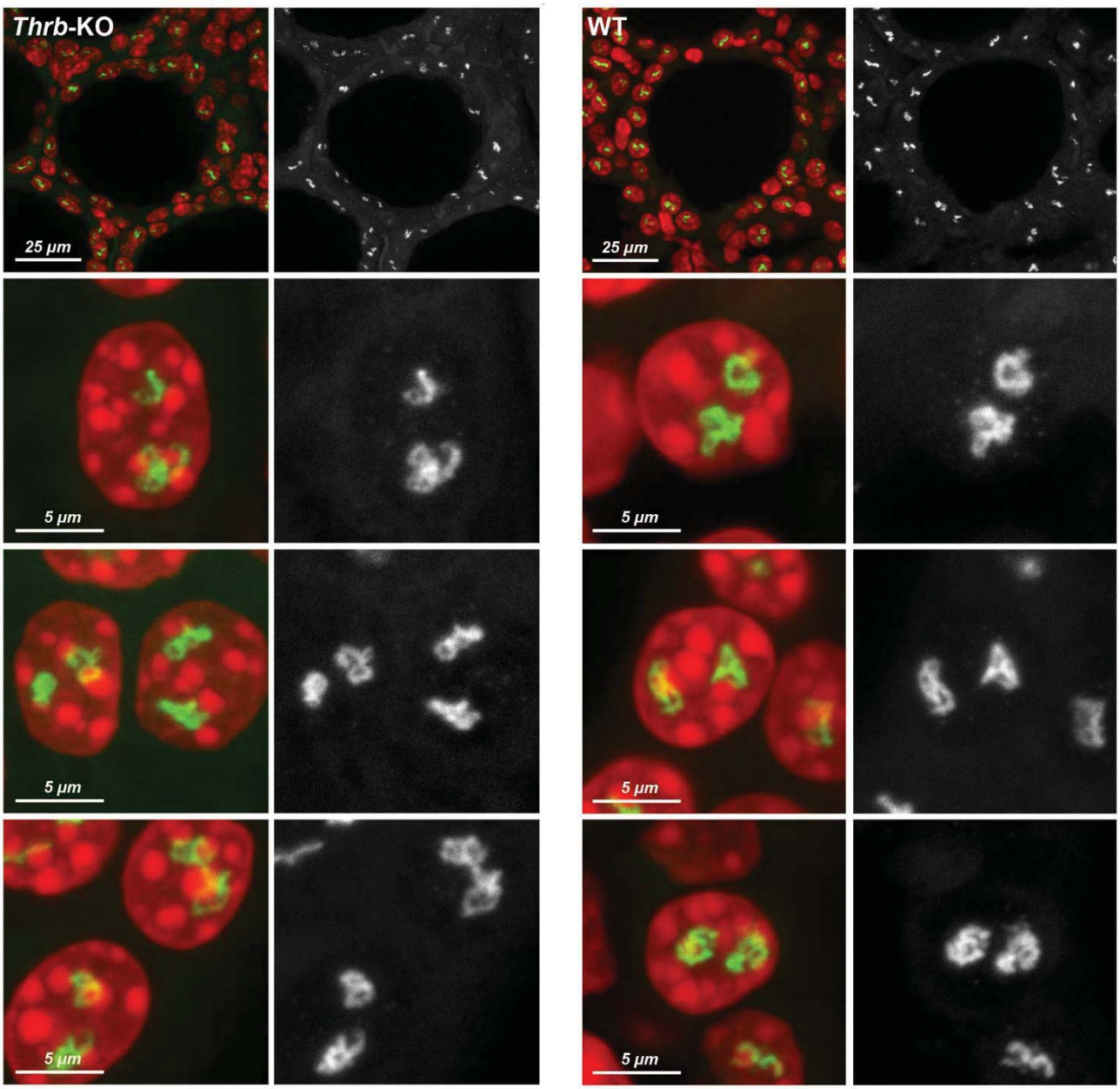


<u>Supplementary Figure 5</u> (appended to FIGURE 5). Follicles and exemplified thyrocyte nuclei after RNA-FISH detecting *Tg* TLs in mice models with decreased (*Trhr1*-KO) and increased (*Mct8*-KO and *Thrb*-KO) thyroidal TH production in comparison to control animals.

***Tg* TLs, *green*; DAPI, *red*; images are projections of 1-3 µm confocal stacks.**

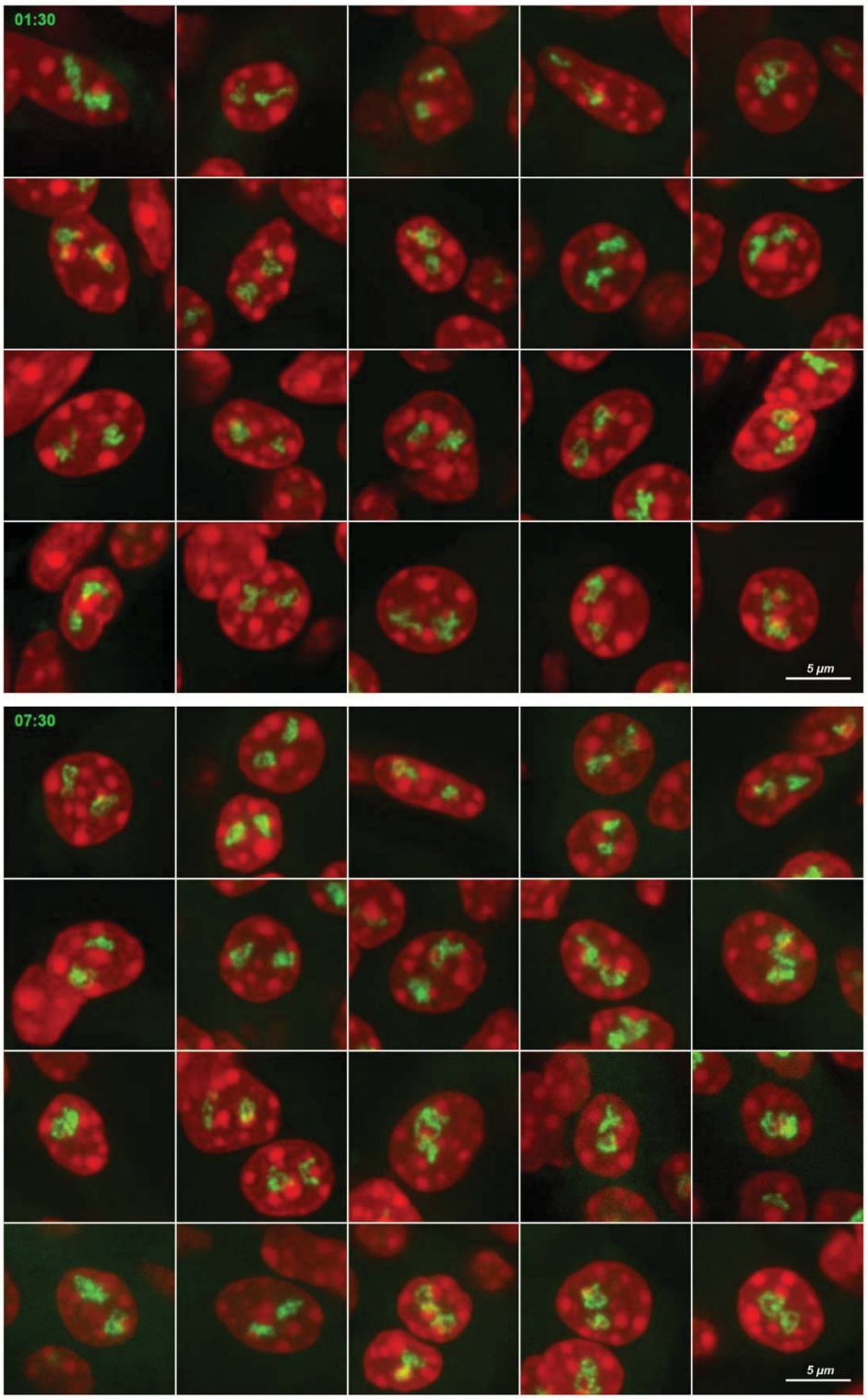


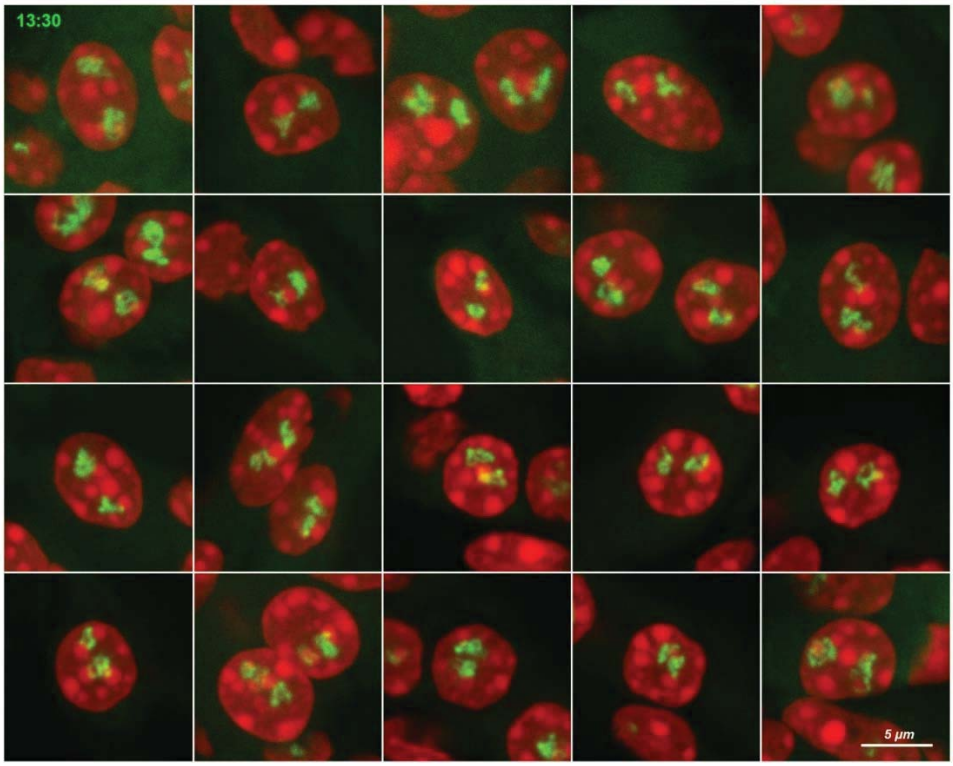

Mct8-KO
WT
25 µm
25 µm
5 µm
5 µm
5 µm
5 µm
5 µm
5 µm

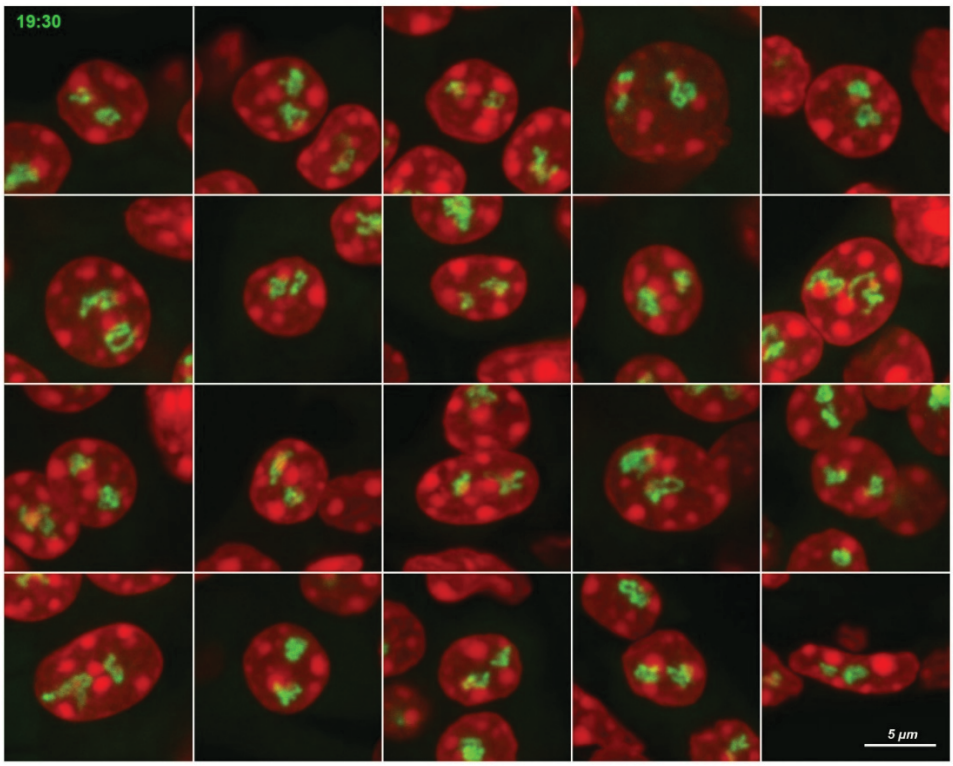

Thrb-KO
WT
25 µm
5 µm

<u>Supplementary Figure 6</u> (appended to FIGURE 6D). 20 examples of thyrocyte nuclei after RNA-FISH detecting *Tg* TLs at the four indicated time points used for the gene circadian activity test.

***Tg* TLs, *green*; DAPI, *red*; images are projections of 3 µm confocal stacks.**




13:30
5 µm


19:30
5 µm

Co-transcriptional splicing is delayed in highly expressed genes

Simon Ullrich*, Iliya Nadelson*, Stefan Krebs, Helmut Blum, Heinrich Leonhardt, Irina Solovei
BioRxiv, 2024 - https://doi.org/10.1101/2024.06.04.597307

Co-transcriptional splicing is delayed in highly expressed genes

Simon Ullrich[1§], Iliya Nadelson[1,2§], Stefan Krebs[3], Helmut Blum[3], Heinrich Leonhardt[1], Irina Solovei[1*]

[1] Faculty of Biology, Ludwig Maximilians University Munich, Planegg-Martinsried, Germany

[2] *Current address:* Institute of Epigenetics and Stem Cells, Helmholtz Zentrum Munich, Germany

[3] Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, Ludwig Maximilians University Munich, Munich, Germany

[*] Corresponded author

[§] These authors contributed equally

ABSTRACT

Despite a significant progress has been made in studies of molecular mechanisms of splicing, the time course of co-transcriptional splicing remains a controversial subject. To elucidate this question further, we made use of long transcription loops formed by the highly upregulated *Tg* gene, microscopy after RNA-FISH and dedicated image analysis. We observed a substantial delay in *Tg* intron splicing - up to 50 kb from a transcription site - which we tentatively explain by the high transcriptional activity of the gene leading to a localized deficiency in the splicing machinery. For comparison, the same assay for the moderately expressed *Cald1* gene revealed no splicing delay supporting our assumption. Phenomenon of the delayed splicing of the highly expressed gene is observed only for the relatively short *Tg* introns with length of 1 -10 kb, whereas the long *Tg* intron (53.8 kb) does not exhibit a measurable splicing delay, suggesting that intron length may serve as a determining factor in modulating the splicing speed. Altogether, our findings provide new insights into splicing of highly transcribed genes and highlight the notion that working with tissues can reveal phenomena not observed in cultured cells.

INTRODUCTION

Eukaryotic genes differ from prokaryotic genes by presence of intragenic non-coding sequences, introns, which play an important role in the regulation of gene expression, gene evolution and protein heterogeneity. This eukaryotic innovation came with a cost of the complicate, energy-consuming and tightly regulated process of intron splicing (de Almeida and Carmo-Fonseca, 2008; Muller-McNicoll and Neugebauer, 2013). Despite a significant progress in splicing studies, the timing of splicing events remains a subject of debates (Eichenberger et al., 2023; Han et al., 2011).

The prevailing classical view on splicing is that it occurs strictly co-transcriptionally (Bentley, 2014; Girard et al., 2012; Martin et al., 2013; Tilgner et al., 2012; Vargas et al., 2011). The spliceosome complex identifies genomic elements defining intron, splice donor, splice acceptor and branchpoint. The complex then excises an entire intron shortly after its transcription, with RNA polymerase II (RNAPII) being 45 nt downstream of the intron 3' border (Oesterreich et al., 2016). In addition, several deviations to the canonical splicing, not mutually exclusive, have

been described. One of them is *post-transcriptional splicing*, when introns are spliced in any order after pre-mRNA polyadenylation (Coté et al., 2024; Gaidatzis et al., 2015). Another variation is *intron retention* (Boutz et al., 2015; Jacob and Smith, 2017), a preservation of one or several specific introns in polyadenylated RNA for the regulation of mRNA release into the cytoplasm, which has been proven to be a widely distributed phenomenon (Braunschweig et al., 2014; Jacob and Smith, 2017). Finally, there is *recursive splicing* when an intron is spliced in multiple steps instead of one (Duff et al., 2015; Sibley et al., 2015) with apparently stochastic splice site selection (Wan et al., 2021).

We have recently shown that highly expressed and long genes expand from their harboring locus and form microscopically resolvable transcription loops (TLs) (Leidescher et al., 2022). The TL extension has been attributed to a dense decoration of a gene with polymerases and attached to them nascent ribonucleoprotein granules (nRNPs), which impose an intrinsic stiffness to the gene axis. In other words, the degree of TL expansion depends on the size of attached nRNPs, which become voluminous over long exons or long introns, and thus depends on splicing dynamics. One of the studied genes, the thyroglobulin gene (*Tg*), is especially highly upregulated and by far exceeds the expression of other genes, including the housekeeping genes (Leidescher et al., 2022; Ullrich et al., 2023). Importantly, although the *Tg* gene is not particularly long (180 kb in mouse), it forms remarkably long TLs extending into nucleoplasm for several microns (Fig. 1A).

The remarkable extension of the *Tg* gene is rather counterintuitive, because the gene has a conventional organization with 48 exons ranging between 66 and 233 nt and introns ranging between 197 and 9,375 kb with only one large intron of approximately 54 kb in length. The *Tg* mRNA is only 8.4 kb long, and providing that splicing occurs strictly co-transcriptionally, the length of nascent RNAs (nRNAs) decorating *Tg* TLs should regularly decrease after each exon, resulting in a small size of nRNPs covering the gene. This, in turn, should allow the gene axis to coil and thus prevent the gene expansion. Nonetheless, TLs of *Tg* expand more than TLs of genes with long exons (e.g., *Ttn*) or long introns (e.g., *Cald1*) (Leidescher et al., 2022). One of the plausible explanations for the remarkable *Tg* TLs extension is a delay of *Tg* intron splicing. Such delay would result in long nRNAs that include not only exons but also several yet-to-be-spliced introns leading to the formation of bulky nRNPs.

The *Tg* gene is expressed exclusively in thyrocytes, the secretory epithelial cells organized into closed follicles within the thyroid gland. When isolated from the tissue and incubated *in vitro*, thyrocytes quickly lose their identity and silence the *Tg* gene (Ullrich et al., 2023). Therefore, the most convenient modern methods of splicing dynamic studies, such as e.g. living cell observations (Martin et al., 2013) or metabolic labeling of nRNAs (Schwalb et al., 2016), are not applicable in this case. Accordingly, to test our hypothesis about delayed splicing of *Tg* nRNAs, we employed several methods compatible with work on thyroid tissue, such as PCR and Nanopore sequencing of RNA isolated from the entire thyroid gland. Furthermore, we performed single cell analysis using oligopainting of consecutive *Tg* introns, followed by robust formal analysis of 3D images. Taken together, the obtained data indicate a substantial delay in splicing of *Tg* introns, which we tentatively explain by local splicing machinery exhaustion due to a high level of the gene transcription.

RESULTS & DISCUSSION

The remarkable length of *Tg* TLs renders them as a unique and attractive model to study transcription and splicing, utilizing light microscopy, as exemplified below with the following three experiments. (I) RNA-FISH with differential labeling of the two halves of the 54 kb long *Tg* intron allows us to track the dynamics of intron splicing. Oligoprobes for the 3'-half label the second half of the intron as expected, while oligoprobes for the 5'-half label the entire intron (Fig. 1B), suggesting that this long intron is excised as a single piece rather than through recursive splicing. (II) RNA-FISH with oligoprobes encompassing 5 kb at the 3'-half of the intron 44 at the 3'-end of *Tg*, demonstrates transcriptional bursting of the gene. Approximately an equal number of nuclei showed oligoprobe signals that were either of a similar size on both alleles, or of different sizes, or present only on one allele (Fig. 1C). Absence of the RNA signals indicates that this particular intron at this particular time remained in a transcription pause, whereas differently sized signals indicate nonsynchronous transcriptional bursting. (III) Finally, our previous studies of TLs formed by long and highly expressed genes demonstrated co-transcriptional splicing on a global scale. Thus, RNA-FISH with two BAC probes highlighting 5' and 3' introns labels the *Tg* TLs sequentially, although with some visible overlapping, because 5' introns become massively spliced out towards the 3' of the TL and thus could not be labeled with the 5' probe (Fig. 1D; see also (Leidescher et al., 2022; Ullrich et al., 2023)). This experiment, however, due to the low resolution of FISH with BAC probes, does not allow an appraisement of detailed splicing dynamics.

RNA signals of *Tg* introns are disproportionally large. To enable the study of splicing dynamics with a greater resolution, we labeled single introns with oligoprobes using SABER-FISH (Kishi et al., 2019). We selected six introns with size varying from 2 to 7 kb located at the 5' end of the gene (Fig. 2A) and three introns located at the 3' end of the Tg gene, including the longest *Tg* intron 40 with the length of 54 kb (Fig. 2B). All single introns were detected by RNA-FISH together with the BAC probe labeling nRNAs decorating 153 kb in the middle of the gene. Unexpectedly, visual pairwise comparison of single intron signals and the BAC signal revealed markedly disproportionally large intron signals in all cases with exception to intron 40 (Fig. 2A,B; Fig. S1A,B).

For the quantification of signal disproportionality, we developed a dedicated workflow for 3D image analysis in confocal stacks. The program automatically separates and segments nuclei, identifies and segments RNA signals, corrects for chromatic shift between channels, and calculates the overlapping volume (V) between BAC and intron signals as a ratio $V_{overlap} / (V_{intron} + V_{BAC})$ (for details see the Methods). Since the RNA signals of two alleles within one nucleus can vary in size due to transcriptional bursting (Fig. 1C), the overlapping volume was estimated for both alleles. The quantitative image analysis confirmed the striking disproportionality of intron signals (Fig. 2C).

Based on the genomic size of the probed regions, an expected overlap proportion of the studied introns must be below 10%. However, the observed overlap ranged between 26% and 49%. For instance, the expected overlap of signal from intron 21 with a length of 7 kb and the BAC encompassing 153 kb of the *Tg* gene, is approximately 4%, but it was measured at above 40% (Fig. 2C). Notably, introns of the same size displayed different mean overlap ratios, whereas

introns of a different length displayed close overlap volumes. As an example, the three introns 21, 23 and 24, with lengths of 7 kb, 2.5 kb and 5.5 kb respectively, all exhibited similar median overlap ratios. Remarkably and somewhat counterintuitively, the longest intron 40 displayed an overlapping value close to the expected one, at around 35% (Fig. 2C).

This disproportionality of RNA intron signals indicated that the probes for introns label not only nRNAs transcribed from the specific intron but also nRNAs in the downstream regions, indicating a delay in the splicing event for 1-10 kb long introns of the *Tg* gene (Fig.3B-D) rather than a strictly co-transcriptional splicing (Fig. 3A).

Splicing of *Tg* introns is delayed for at least 0.75 kb. To confirm the occurrence of a splicing delay, we extracted total RNA from mouse thyroids and carried out PCR with primers designed for regions approximately 250 bp upstream of an exon 5'-border and 750 bp downstream of its 3'-border. For all four tested exons, the intron-exon-intron sequences were successfully amplified, confirming that splicing of these introns is retarded for at least 0.75 kb (Fig. S2). This finding aligns with previous studies showing that in some human genes, fully transcribed introns may remain within nRNAs for up to 2 kb downstream of the splice junction (Drexler et al., 2020; Sousa-Luis et al., 2021). However, the degree of Tg intron signal disproportionality suggests that splicing delay during *Tg* transcription is greater than several kb. Therefore, we attempted to detect *Tg* nRNAs at a high-throughput level to examine longer fragments of nascent transcripts.

Nanopore RNA sequencing demonstrates massive retardation of *Tg* intron in nRNAs. Detection of delayed splicing or intron retention from bulk RNA-seq data is a problematic issue primarily due to the intrinsically short RNA-seq reads not sufficient for characterizing single or multiple introns within an individual transcript. Recently, several highly efficient parallel methods for sequencing of nRNA have emerged, such as GRO-seq (Core et al., 2008), NET-seq (Mayer and Churchman, 2016), TT-Seq (Schwalb et al., 2016) and POINT technology (Sousa-Luis et al., 2021). These techniques demand either metabolic labeling or the isolation of chromatin associated RNA. However, these requirements pose great challenges for cells in tissues, especially if a tissue is inherently difficult to dissociate or cultivate primary cells from, while preserving the transcriptional program – both are the case for the thyroid gland (Ullrich et al., 2023).

Therefore, to elucidate the dynamics of delayed splicing events, we adopted an unconventional Nanopore sequencing workflow through enriching the sequencing library for nRNAs by poly(A)-depletion of the total RNA isolated from the tissue. The length distribution of *Tg* RNA intron-containing reads indicates an even sequencing coverage of nascent *Tg* RNAs (Fig. 4A, grey columns). For the analysis, only intron-containing reads were considered. Multiple reads contained one or more fully transcribed introns, while some included only partial intron regions, thus not allowing to differentiate between introns that had already been transcribed and those still in the process of transcription. The read-length spanned up to 8.5 kb, which is equivalent to the size of fully spliced and mature mRNA, however, a fraction of reads exceeded the size of 8.5 kb and reached up to 15 kb (Fig. 4A, blue columns), doubtlessly indicating the presence of unspliced introns. The coverage across all intronic regions within the *Tg* gene was rather comprehensive and uniform (Fig. 4B), indicative of the robustness of our approach in capturing nascent transcripts.

Unsurprisingly, the distribution analysis of intron-containing reads showed a predominant presence of reads containing only a single intron. Nonetheless, a discernible fraction of reads displayed the incorporation of multiple introns within a single read (Fig. 4C) including up to eight introns. Our previous data on Nanopore sequencing of the poly(A)-enriched mRNA fraction extracted from mouse thyroid tissue showed that there is no intron retention in the poly(A) *Tg* mRNA fraction (Ullrich et al., 2023). Therefore, we could rule out that the observed splicing retardation in our reads was due to regulatory intron retention for controlled mRNA export into the cytoplasm. Consequently, we concluded that our data indeed reflects delayed splicing.

Furthermore, the examination of individual reads unveiled instances in which delayed splicing was unmistakably manifested by the presence of multiple unspliced introns in a single read, sometimes in combination with already spliced introns (Fig. 4D). Thus, collectively, our data revealed a delay in splicing across both single and multiple introns within individual *Tg* transcripts. However, due to technical limitations, they do not indicate how long the delay in splicing for individual introns can be.

Although the Nanopore sequencing allows long-read direct RNA sequencing of nascent transcripts (Garalde et al., 2018) and has been used as a powerful tool for dissecting the intricacies of RNA processing dynamics (Parker et al., 2020; Sousa-Luis et al., 2021), in case of the thyroid tissue, we cannot draw reliable conclusions, which we tentatively attribute to the small amount of used RNA. According to the manufacturer, to obtain accurate and reliable results using this method, the amount of input RNA should be in the range of 1000 ng of total RNA. Given that thyrocytes comprise only 60% of the thyroid gland and that a single mouse thyroid yields less than 200 ng of total RNA, of which only a tiny proportion represents nRNAs, the necessary quantity of thyroids and sacrificed mice for one experiment exceed a reasonable number. Therefore, we searched for another approach to estimate a splicing delay magnitude.

Pairwise comparison of intron signals confirms splicing delay. We reasoned that the considerable length of *Tg* TLs (Fig. 1A) allows for the examination of splicing on a single cell level using FISH and microscopy. Firstly, since RNA-FISH does not require the denaturation of DNA in sections, probes for introns do not hybridize to the gene DNA and reliably detect nRNAs. Secondly, the visualization of differentially labeled introns in a pairwise manner would allow a signal comparison of a reference intron with downstream introns (Fig. 3E). We expected that if there is a delay in splicing of an intron, its signal would colocalize with a signal from the subsequent intron or even multiple introns. Given our observation of co-transcriptional splicing on the *Tg* gene on a global scale (Fig. 1D), we expected to observe a separation of the intron signals as the genomic distance between them increases (Fig. 3E).

For such analysis, we used the same regions in the beginning and the end of the gene as in the previous RNA-FISH experiment (Fig. 2, 5). Within each region, we selected a reference intron at the 5' end - specifically, introns 19 and 38 - and compared their signals with subsequent downstream test introns. For each comparison, we estimated the colocalization of introns by the Pearson's correlation coefficient (PCC). Differentially labeled nRNAs of the reference and test intron were detected in two spectrally separated channels, allowing pixel intensity based colocalization analysis. Such analysis includes the evaluation of the intensity of each pixel in one channel against the intensity of the corresponding pixel in another channel, generating a correlation coefficient (Cordelieres and Bolte, 2014; Cordelières and Zhang, 2020). In addition,

we set up positive and negative controls for both regions. For the positive control demonstrating maximally positive correlation, we used the same reference intron labeled with two distinct fluorophores. For the negative control, we compared the reference and a downstream intron positioned at a minimum distance of 50 kb apart.

We modified our analysis pipeline to calculate the PCC between signals in two channels within a 3D volume (for details see Methods). Importantly, the collected image stacks underwent correction for both axial and lateral chromatic shifts between spectrally separated channels. To achieve this, stacks of 0.5 µm beads labeled with four fluorophores were collected during each acquisition session and processed alongside the samples.

As expected, signals of both fluorophores over the reference introns visually overlapped (green fields in Fig. 5A,B) and the mean PCC for the positive controls was close to 1 (Fig. 4C). On the contrary, the signals of negative controls were visually separated (pink fields in Fig. 5A,B) and accordingly their PCC values were negative (Fig. 5C). It is important to note that the anti-correlation does not hold any specific biological meaning but rather indicates a separation of the signals. The mean PCC values for 7 pairwise comparisons between reference and tested introns ranged between the values for positive and negative controls, consistently decreasing as the genomic distances between them increased (Fig. 5C), corresponding to visual estimation (Fig. 5A,B; Fig. S3A,B).

Based on these results, we conclude that intron splicing might be delayed for up to 50 kb. Admittedly, our microscopy analysis deals with structures below 0.5-1 µm that are close to the resolution limit of light microscopy. Therefore, we cannot fully exclude the possibility that the colocalization of intron signals, to some degree, is influenced by the coiling of the gene axis, leading to an abutting of neighboring introns and thus mixing of their nRNA signals. However, our observations on the disproportionality of single intron signals in comparison to an almost entire TL (Fig. 2) argue that axis coiling alone cannot entirely explain the observed phenomenon.

Splicing delay is connected to a high transcription level. The *Tg* gene is characterized by an exceptionally high expression level of about 23,000 TPM (Leidescher et al., 2022). We hypothesize that such a high transcription level might cause a localized depletion of splicing machinery, hence resulting in splicing delays. Unfortunately, we cannot directly test this hypothesis by analyzing other genes with comparable length and high expression levels, as we are unaware of any other mammalian genes meeting these criteria. Conversely, we used a comparably long gene but with a moderate expression, namely the *Cald1* gene. *Cald1* has a length of 177 kb, is expressed in cultured myoblasts at approximately 1,800 TPM and correspondingly forms small but microscopically resolvable TLs (Leidescher et al., 2022). Similarly to the *Tg* gene, we tested splicing dynamics of the *Cald1* by the pairwise colocalization comparison of the intron signals (Fig. 6).

The first Cald1 intron was used as a reference and colocalization of its signal was measured pairwise with the four subsequent test introns (Fig. 6A). By visual inspection of confocal stacks, only the signals from the reference intron labeled with two fluorophores exhibited colocalization, whereas signals for the subsequent introns, although tightly adjacent, were distinguishable from the reference (Fig. 6B). Consequently, compared to the reference intron, the rest of the comparisons yielded a PCC value at zero or below (Fig. 6C). Therefore, this data supports our conclusion that global splicing dynamics are regulated by the level of transcription.

Splicing dynamics for short and long introns is different. The proportional size of the *Tg* intron 40 signal compared to the BAC signal revealed in the initial experiment (Fig. 2B,C), suggests that there is no delay in the splicing of this particular intron. To investigate this further, we designed another experiment, in which three sequential introns – 40 (*53.8 kb*), 41 (*1 kb*) and 42 (*9.4 kb*) - are labeled with three different fluorophores (Fig. 7A1). This setup allowed us to perform three pairwise comparisons of the introns (Fig. 7A2). In agreement with the previous experiment revealing a delay in splicing of *Tg* introns (Fig. 5), the two sequential introns, 41 and 42, demonstrated a high degree of colocalization. Surprisingly, the degree of colocalization between introns 40 and 41 was significantly lower and became negative when assessed for introns 40 and 42 (Fig. 7A3). On the contrary, three short sequential introns – 22 (1.9 kb), 23 (2.5 kb), 24 (5.3 kb) (Fig. 7B1) - labeled with three different fluorophores yielded high PCCs in all pairwise comparisons, as expected (Fig. 7B2, B3).

The introns 40 and 42 are separated by only 10.4 kb and thus their negative correlation is in conflict with the data on colocalization of 19-21 and 19-22 introns, which are also separated by 10-12 kb, but exhibit a PCC of around 0.5 (Fig. 5C). This odd behavior was observed solely for the longest *Tg* intron 40, spanning an overall length of 53.8 kb, indicating that intron length could indeed be an influencing factor for the splicing speed.

The phenomenon of a quicker splicing of the long compared to the short introns, especially considering the lack of recursive splicing (Fig. 1B), appears to be counterintuitive. We can propose three hypothetical scenarios to explain the phenomenon. Firstly, a reduced speed of RNAPIIs over long introns could provide more time for the assembly of the splicing machinery. However, this explanation is highly speculative as RNAPII elongation speed is generally considered constant and estimated at approximately 3.8 kb/min (Singh and Padgett, 2009). The second scenario considers the pronounced extension of intron 40 from the entire *Tg* TL (Fig. 2B; Fig. S1B; (Leidescher et al., 2022)), which brings the intron into a less competitive environment for the splicing machinery. And finally, providing that the transcription speed is 3.8 kb/min, transcription of the entire 53.8 kb intron would take more than 14 min, allowing more time to assemble the necessary splicing machinery compared to short introns. The last explanation seems to be most probable and has been recently proposed as an explanation of positive correlation between splicing speed and intron length (Sousa-Luis et al., 2021).

CONCLUSIONS

Our data indicate that there is a substantial delay in intron splicing, up to 50 kb, during transcription of the highly upregulated *Tg* gene, which is not observed for the moderately expressed *Cald1* gene. We tentatively link the splicing delay to the high transcriptional activity of the gene, potentially leading to a localized deficiency in the splicing machinery. Intriguingly, the long *Tg* intron stands out as an exception from this phenomenon and does not exhibit a measurable splicing delay.

Splicing delay can occur according to different scenarios: (1) only certain introns are spliced with delay (Fig. 3B), (2) intron sequences persist across several exons and then undergo massive splicing at a particular time point (Fig. 3C), (3) in only a fraction of nRNAs, introns undergo asynchronous splicing downstream of their transcription sites (Fig. 3D). Most likely, a

combinations of (1) and (3), or (2) and (3) scenarios co-exist during an intense transcription, however, the resolution of our study at both microscopy and Nanopore sequencing levels, does not allow to distinguish between these three possible scenarios.

Irrespective of the possible scenarios discussed above, the delayed splicing explains why *Tg* TLs extend so much into the nuclear interior. Due to an accumulation of introns in nRNAs, the size of nRNPs gradually increases over tens of kilobases. The resulting nRNP bulkiness in turn increases gene axis stiffness and allows the gene to extend into the nuclear space.

Altogether, our findings shed light on the intricate and dynamic nature of RNA processing, hinting at underlying complexities in the regulation of splicing events during the transcript maturation. In addition, our work emphasizes the power of microscopy and image analysis in studies of transcription and demonstrates that working with tissues can be highly rewarding. Despite apparent restrictions and technical difficulties, tissues offer phenomena and processes *in vivo*, which might not be present in cells cultured *in vitro*.

MATERIAL AND METHODS

Tissue collection and cryosections. Thyroid tissue sampling was executed in accordance with the European Union (EU) directive 2010/63/EU on the protection of animals used for scientific purposes and in compliance with regulations by the respective local Animal Welfare Committees (LMU; Committee on Animal Health and Care of the local governmental body of the state of Upper Bavaria; Germany). Mice were sacrificed by cervical dislocation after IsoFlo (Isofluran, Abbott) narcosis. Freshly dissected tissues were washed with PBS and then fixed with 4% paraformaldehyde (Carl Roth) solution in PBS for 12-20 h. After fixation, thyroids were washed with PBS, cryoprotected in a series of sucrose, and embedded in Tissue-Tek O.C.T. compound freezing medium (Sakura). Blocks were stored at - 80°C before cutting into 16-20 µm sections using a cryostat (Leica CM3050S). Cryosections were collected on Superfrost Plus slides (Thermo Scientific) and stored at - 80°C before use.

Cell culture. The 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. Cells were sub-cloned on coverslips pretreated with polylysine. After brief wash with pre-warmed PBS, cells were fixed with 4% paraformaldehyde (Carl Roth) solution in PBS for 10 min, washed with PBS (10 min x 3), permeabilized with 0.5% Triton X100 in PBS, washed with PBS supplemented with 0.01% Tween 20 (10 min x 3), equilibrated in 2xSSC and stored in 50% formamide in 2xSSC buffer.

Probes. BAC clones encompassing the *Tg* gene (RP24-229C15, RP23-193A18, RP23-266I10) were selected using the UCSC genome browser and purchased from BACPAC Resources (Oakland children's hospital) as agar stabs (https://bacpacresources.org/). BACs were purified via standard alkaline lysis or the NucleoBond Xtra Midi Kit (Macherey-Nagel), followed by amplification with the GenomiPhi Kit (GE Healthcare) according to the manufacturer's instructions. Amplified BAC DNA was labeled with fluorophores using homemade conjugated fluorophore-dUTPs by nick translation (Cremer et al., 2008). Labeled BAC DNA was ethanol precipitated with 10-fold excess of Cot-1 (1 mg/ml; Invitrogen, 18440-016) and 50-fold excess of salmon sperm DNA (5 µg/µl; Sigma); pellet was dried in a SpeedVak, and dissolved in hybridization mixture containing 50% formamide, 1xSSC and 10% of dextran sulfate.

Genomic coordinates of the studied introns are listed in Supplementary Table 1. Oligoprobes for introns were generated using SABER-FISH protocol and described in detail previously (Leidescher et al., 2022). Briefly, oligos targeting intron sequences were designed using Paintshop (Hershberg et al., 2021) and ordered as oligo pools from Integrated DNA Technologies. The oligos were remapped if necessary for multi-colour imaging and extended to ~500 nt using primer exchange reaction (Kishi et al., 2019). Finally, the probes were purified using PCR clean-up columns (Macherey-Nagel).

FISH. RNA-FISH using BACs was performed on cryosections and cultured myoblasts as previously described. Denatured probes were loaded on sections under small glass chambers or on coverslips with myoblasts and sealed with rubber cement (for detail, see (Eberhart et al., 2012; Solovei and Cremer, 2010). Denaturation of sections and cells was omitted. Hybridizations were carried out in a water bath at 37°C for 1-2 days. After hybridization, rubber cement and chambers were removed, slides with sections or coverslips with myoblasts were washed with 2xSSC at 37°C, 3 x 30 min, and then with 0.1xSSC at 60°C 1 x 7 min. Hybridized SABER probes were detected by incubation with 1 µM fluorescently labeled detection oligonucleotides in PBS for 1 h at 37°C followed by washing with PBS for 10 min.

Microscopy. Confocal 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. Chromatic shift was measured using Tetraspeck beads (0.5 µm).

Image Analysis. All image sets (i.e., z-stack imaged in three spectrally separated channels) for an overlap or correlation experiment were processed automatically with custom python (v3.8.0; Python Software Foundation) scripts, making use of several packages: numpy (Harris et al., 2020), pandas (McKinney, 2011), skimage (Van der Walt et al., 2014), scipy (Virtanen et al., 2020), matplotlib (Hunter, 2007), seaborn (Waskom, 2021), tifffile (Gohlke) and napari (Ahlers et al.). The scripts are available upon request. The main steps are summarized below.

Calculation of overlapping of FISH signals in 3D. First, each image set was corrected for chromatic shift between the channels. Next, the DAPI signal was binarised using Otsu's method (Otsu, 1975) and cleaned from small artefacts. Subsequently, a watershedding procedure was employed to separate the nucleus in the centre of the image from neighbouring nuclei. Objects touching the image borders were removed and the object with the biggest volume was retained as the segmented nucleus.

Within the segmented nucleus, the signal of each FISH channel was segmented. First, the volume around each FISH signal was reduced by calculating and thresholding (Otsu) a local entropy image. The resulting mask was applied to the original FISH image, which was then also thresholded using Otsu's method. The number of remaining objects was evaluated and a maximum of two objects (biggest volumes) per FISH channel was retained.

Finally, the overlap between objects in both FISH channels was calculated and compared to the total volume of objects in both FISH channels. Specifically, the signals of single introns were overlapped with a BAC probe delineating most of the gene. To ensure only results from correctly segmented signals remained, image sets where less than 80% of the segmented intron overlapped with the segmented BAC were excluded from the analysis.

Calculation of correlation between FISH Signals in 3D. For the pixel-based colocalization analysis of FISH signals, images were processed as described above until the FISH segmentation. The volume around the FISH signals was further confined by summing both FISH channels and thresholding the result (Otsu's method) while applying the nuclear mask. This reduced the influence of background fluorescence on colocalization calculations. The resulting mask was cleaned-up from artefacts and objects overlapping with the nuclear border were removed. The FISH signals were smoothed (Gaussian filter, σ=1) before the mask was applied to reduce the value differences between neighbouring pixels. Finally, the pixel-wise Pearson Correlation Coefficient (PCC) between the resulting volumes of both FISH signals was computed. The resulting correlation values were saved in .csv format and plotted using R Statistical Software.

PCR analysis. Dissected thyroids were immediately placed in Lysis Buffer RA1 and homogenized using an ultra-turrax dispersing tool. Total RNA was isolated using the NucleoSpin RNA Kit (Macherey-Nagel). 1 µg of total RNA was reverse transcribed using Maxima H Minus Reverse Transcriptase (Thermo Scientific) with gene specific primers. PCR was performed in technical and biological triplicates using Phusion Plus polymerase (Thermo Scientific) and subsequently analysed on 1% agarose gel. All steps were performed according to the manufacturer's recommendation. Distances between introns were calculated from the 3' end of a reference intron to the 3' end of a test intron. All used primers can be found in Supplementary Table 2.

Nanopore sequencing. Freshly dissected thyroids were immediately placed in ice cold TriZol and homogenized using an ultra-turrax dispersing tool. Poly(A)+ transcripts were isolated using magnetic oligoT-beads (Lexogen). The sequencing library was created using the PCR-cDNA Sequencing Kit (PCB111.24, Oxford Nanopore) and sequenced on a PromethION P24 on a R9.4.1 flowcell. The sequencing data were basecalled using Guppy v6.4.6 and mapped to the mouse genome (mm10) using minimap2 (Li, 2018).

ACKNOWLEDGEMENTS

We are grateful to Maria Carmo-Fonseca (Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina da Universidade de Lisboa) and Daniel Larson (Receptor Biology and Gene Expression, NIH) for fruitful discussions.

COMPETING INTERESTS

The authors declare no competing or financial interests.

AUTHOR CONTRIBUTIONS

Conceptualization: I.S.; *Methodology:* S.U., I.N., I.S.; *Validation:* S.U., I.N., I.S.; *Formal analysis:* S.U., I.N.; *Investigation*. S.U., I.N., I.S., S.K., H.B.; *Resources:* S.K., H.B.; *Data curation:* I.S., H.L.; *Writing - original draft:* S.U., I.S., I.N.; *Writing - review & editing*: S.U., I.S., I.N., H.L.; *Visualization*: S.U., I.S., I.N.; *Supervision:* I.S.; *Project administration*: I.S.; *Funding acquisition:* I.S., H.L.

FUNDING

This work has been supported by the Deutsche Forschungsgemeinschaft grants (SP2202/SO1054/2, project # 422388934 to IS, SPP 2202/LE721/17-1, project # 422857584 to HL, and SFB1064, project # 213249687 to HL and IS)

DATA AVAILABILITY

All relevant data can be found within the article and its supplementary information. Data on Nanopore sequencing are deposited to SRA database with PRJNA1112134 accession number.

REFERENCES

Ahlers, J., D. Althviz Moré, O. Amsalem, A. Anderson, G. Bokota, P. Boone, J. Bragantini, G. Buckley, A. Burt, M. Bussonnier, A. Can Solak, C. Caporal, D. Doncila Pop, K. Evans, J. Freeman, L. Gaifas, C. Gohlke, K. Gunalan, H. Har-Gil, M. Harfouche, K.I.S. Harrington, V. Hilsenstein, K. Hutchings, T. Lambert, J. Lauer, G. Lichtner, Z. Liu, L. Liu, A. Lowe, L. Marconato, S. Martin, A. McGovern, L. Migas, N. Miller, H. Muñoz, J.-H. Müller, C. Nauroth-Kreß, J. Nunez-Iglesias, C. Pape, K. Pevey, G. Peña-Castellanos, A. Pierré, J. Rodríguez-Guerra, D. Ross, L. Royer, C.T. Russell, G. Selzer, P. Smith, P. Sobolewski, K. Sofiiuk, N. Sofroniew, D. Stansby, A. Sweet, W.-M. Vierdag, P. Wadhwa, M. Weber Mendonça, J. Windhager, P. Winston, and K. Yamauchi. napari: a multi-dimensional image viewer for Python. Zenodo.

Bentley, D.L. 2014. Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*. 15:163-175.

Boutz, P.L., A. Bhutkar, and P.A. Sharp. 2015. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes & development*. 29:63-80.

Braunschweig, U., N.L. Barbosa-Morais, Q. Pan, E.N. Nachman, B. Alipanahi, T. Gonatopoulos-Pournatzis, B. Frey, M. Irimia, and B.J. Blencowe. 2014. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res*. 24:1774-1786.

Cordelieres, F.P., and S. Bolte. 2014. Experimenters' guide to colocalization studies: finding a way through indicators and quantifiers, in practice. *Methods in cell biology*. 123:395-408.

Cordelières, F.P., and C. Zhang. 2020. 3D quantitative colocalisation analysis. *Bioimage Data Analysis Workflows*. 969:33-66.

Core, L.J., J.J. Waterfall, and J.T. Lis. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 322:1845-1848.

Coté, A., A. O'Farrell, I. Dardani, M. Dunagin, C. Coté, Y. Wan, S. Bayatpour, H.L. Drexler, K.A. Alexander, and F. Chen. 2024. Post-transcriptional splicing can occur in a slow-moving zone around the gene. *Elife*. 12:RP91357.

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

de Almeida, S.F., and M. Carmo-Fonseca. 2008. The CTD role in cotranscriptional RNA processing and surveillance. *FEBS letters*. 582:1971-1976.

Drexler, H.L., K. Choquet, and L.S. Churchman. 2020. Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Molecular cell*. 77:985-998. e988.

Duff, M.O., S. Olson, X. Wei, S.C. Garrett, A. Osman, M. Bolisetty, A. Plocik, S.E. Celniker, and B.R. Graveley. 2015. Genome-wide identification of zero nucleotide recursive splicing in Drosophila. *Nature*. 521:376-379.

Eberhart, A., H. Kimura, H. Leonhardt, B. Joffe, and I. Solovei. 2012. Reliable detection of epigenetic histone marks and nuclear proteins in tissue cryosections. *Chromosome research*. 20:849-858.

Eichenberger, B.T., E. Griesbach, J. Mitchell, and J.A. Chao. 2023. Following the Birth, Life, and Death of mRNAs in Single Cells. *Annual Review of Cell and Developmental Biology*. 39:253-275.

Gaidatzis, D., L. Burger, M. Florescu, and M.B. Stadler. 2015. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nature biotechnology*. 33:722-729.

Garalde, D.R., E.A. Snell, D. Jachimowicz, B. Sipos, J.H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James, and A. Warland. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nature methods*. 15:201-206.

Girard, C., C.L. Will, J. Peng, E.M. Makarov, B. Kastner, I. Lemm, H. Urlaub, K. Hartmuth, and R. Lührmann. 2012. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nature communications*. 3:994.

Gohlke, C. cgohlke/tifffile: v2023.7.4. Zenodo.

Han, J., J. Xiong, D. Wang, and X.-D. Fu. 2011. Pre-mRNA splicing: where and when in the nucleus. *Trends in cell biology*. 21:336-343.

Harris, C.R., K.J. Millman, S.J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, and N.J. Smith. 2020. Array programming with NumPy. *Nature*. 585:357-362.

Hershberg, E.A., C.K. Camplisson, J.L. Close, S. Attar, R. Chern, Y. Liu, S. Akilesh, P.R. Nicovich, and B.J. Beliveau. 2021. PaintSHOP enables the interactive design of transcriptome-and genome-scale oligonucleotide FISH experiments. *Nature methods*. 18:937-944.

Hunter, J.D. 2007. Matplotlib: A 2D graphics environment. *Computing in science & engineering*. 9:90-95.

Jacob, A.G., and C.W. Smith. 2017. Intron retention as a component of regulated gene expression programs. *Human genetics*. 136:1043-1057.

Kishi, J.Y., S.W. Lapan, B.J. Beliveau, E.R. West, A. Zhu, H.M. Sasaki, S.K. Saka, Y. Wang, C.L. Cepko, and P. Yin. 2019. SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nature methods*. 16:533-544.

Leidescher, S., J. Ribisel, S. Ullrich, Y. Feodorova, E. Hildebrand, A. Galitsyna, S. Bultmann, S. Link, K. Thanisch, and C. Mulholland. 2022. Spatial organization of transcribed eukaryotic genes. *Nature cell biology*. 24:327-339.

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 34:3094-3100.

Martin, R.M., J. Rino, C. Carvalho, T. Kirchhausen, and M. Carmo-Fonseca. 2013. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell reports*. 4:1144-1155.

Mayer, A., and L.S. Churchman. 2016. Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. *Nature protocols*. 11:813-833.

McKinney, W. 2011. pandas: a foundational Python library for data analysis and statistics. *Python for high performance and scientific computing*. 14:1-9.

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

Oesterreich, F.C., L. Herzel, K. Straube, K. Hujer, J. Howard, and K.M. Neugebauer. 2016. Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell*. 165:372-381.

Otsu, N. 1975. A threshold selection method from gray-level histograms. *Automatica*. 11:23-27.

Parker, M.T., K. Knop, A.V. Sherwood, N.J. Schurch, K. Mackinnon, P.D. Gould, A.J. Hall, G.J. Barton, and G.G. Simpson. 2020. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m6A modification. *Elife*. 9:e49658.

Schwalb, B., M. Michel, B. Zacher, K. Frühauf, C. Demel, A. Tresch, J. Gagneur, and P. Cramer. 2016. TT-seq maps the human transient transcriptome. *Science*. 352:1225-1228.

Sibley, C.R., W. Emmett, L. Blazquez, A. Faro, N. Haberman, M. Briese, D. Trabzuni, M. Ryten, M.E. Weale, J. Hardy, M. Modic, T. Curk, S.W. Wilson, V. Plagnol, and J. Ule. 2015. Recursive splicing in long vertebrate genes. *Nature*. 521:371-375.

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

Solovei, I., and M. Cremer. 2010. 3D-FISH on cultured cells combined with immunostaining. *Methods Mol Biol*. 659:117-126.

Sousa-Luis, R., G. Dujardin, I. Zukher, H. Kimura, C. Weldon, M. Carmo-Fonseca, N.J. Proudfoot, and T. Nojima. 2021. POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Molecular cell*. 81:1935-1950. e1936.

Tilgner, H., D.G. Knowles, R. Johnson, C.A. Davis, S. Chakrabortty, S. Djebali, J. Curado, M. Snyder, T.R. Gingeras, and R. Guigó. 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome research*. 22:1616-1625.

Ullrich, S., S. Leidescher, Y. Feodorova, K. Thanisch, J.-B. Fini, B. Kaspers, F. Weber, B. Markova, D. Führer, and M. Romitti. 2023. The highly and perpetually upregulated thyroglobulin gene is a hallmark of functional thyrocytes. *Frontiers in Cell and Developmental Biology*. 11.

Van der Walt, S., J.L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner, N. Yager, E. Gouillart, and T. Yu. 2014. scikit-image: image processing in Python. *PeerJ*. 2:e453.

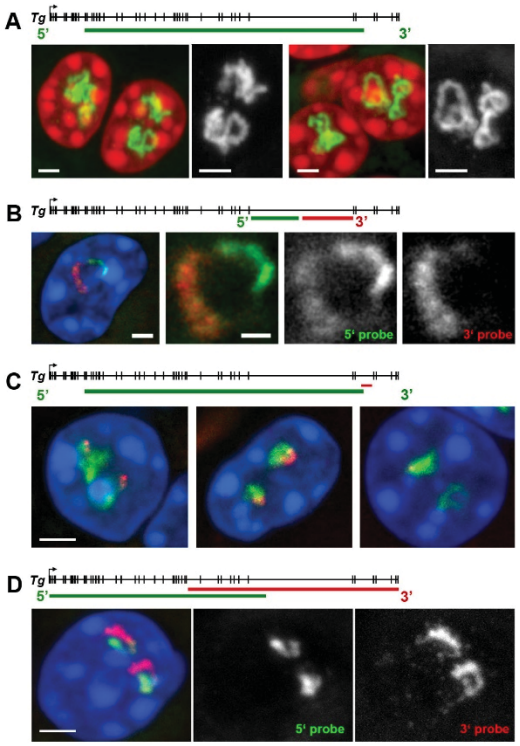
Vargas, D.Y., K. Shah, M. Batish, M. Levandoski, S. Sinha, S.A. Marras, P. Schedl, and S. Tyagi. 2011. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell*. 147:1054-1065.

Virtanen, P., R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, and J. Bright. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*. 17:261-272.

Wan, Y., D.G. Anastasakis, J. Rodriguez, M. Palangat, P. Gudla, G. Zaki, M. Tandon, G. Pegoraro, C.C. Chow, and M. Hafner. 2021. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell*. 184:2878-2895. e2820.

Waskom, M.L. 2021. Seaborn: statistical data visualization. *Journal of Open Source Software*. 6:3021.

Fig. 1. Strong extension of the *Tg* TLs allows transcription analysis using light microscopy.



A, several examples of *Tg* TLs (*green*) within thyrocyte nuclei (*red*) detected by RNA-FISH using genomic BAC probe encompassing ca. 150 kb of the 180 kb long *Tg*. Note the volatile and strongly convoluted shape of TLs. **B,** labelling of intron 40 by RNA-FISH with oligoprobes for 5' (*green*) and 3' (*red*) halves of the intron. Note that the 5' probe labels the entire intron, whereas the 3' probe labels only half of the intron. The gradient in the nRNAs reflects the direction of transcription. **C,** the oligoprobe for the intron 44 at the 3' end of the gene (*red*) in combination with genomic probe (*green*) after RNA-FISH. Three examples demonstrate similar intron signals (*left*), different sizes (*mid*) and signal on only one of the alleles (*right*). **D,** labelling of the *Tg* TLs with two BAC clones encompassing 5' (*green*) and 3' (*red*) halves of the loops by RNA-FISH. Although the probes overlap in the middle of the gene, on a global scale, they label the TLs sequentially as a result of co-transcriptional splicing.

Every panel is accompanied by schematics (above images) illustrating the coverage of the gene with probes. Images are projections of short confocal stacks; nuclei are counterstained with DAPI. RGB panels are supplemented with grey scale images of the RNA signals. Scale bars: A,C,D, *2 µm*; B, *1 µm* and *0.5 µm* in close-ups.

Fig. 2. RNA signals of single *Tg* introns are disproportionally large in comparison to RNA signal of a genomic probe.

A, B, pairwise comparison of RNA signals of single intron *versus* BAC probe (153 kb) probe in the 5' (A, *light-grey rectangle*) and 3' (B, *dark-grey rectangle*) gene regions. The green lines indicate the BAC probe; on the close-ups of the sampled regions, the *arrowheads* mark the 5'-end of introns, for which oligoprobes were designed.

Exemplifying images of thyrocytes after RNA-FISH with pairwise probes (BAC, *green*; introns, *red*) are shown below the schematics. For clarity, RGB images are supplemented with grey scale images of RNA signals. Numbers and lengths of introns are indicated on the grey scale intron panels. Scale bars: *2 µm*. For more examples of nuclei in this experiment, see Fig. S1.

C, boxplots showing the overlap ratio of intron signals with the BAC probe from both gene regions. Introns are arranged on the X-axis from smallest to largest; their numbers indicated below and their lengths - above the X-axis. Thick lines within boxes indicate the medians; boxes show 25th and 75th percentiles; whiskers represent the smallest and the largest values within 1.5 times the interquartile range; grey dots are individual data points. The number of evaluated nuclei ranged between 23-27.

Note that the signal overlap of similarly sized introns can be significantly different, as it is the case for intron 20 and intron 23, both 2.5 kb in size ($p < 0.001$). On the contrary, introns that differ in their size, such as intron 23 (2.5 kb) and intron 24 (5.5 kb), can show similar overlap ratios ($p > 0.05$, one-way ANOVA with Tukey's multiple comparisons post hoc). Expected overlap proportions based on the genomic size of the probes are marked with red lines. Note that only in case of the long intron 40, expected and measured overlaps are similar; for all other introns, the measured overlap exceeded the expected value 6- to 10-fold.

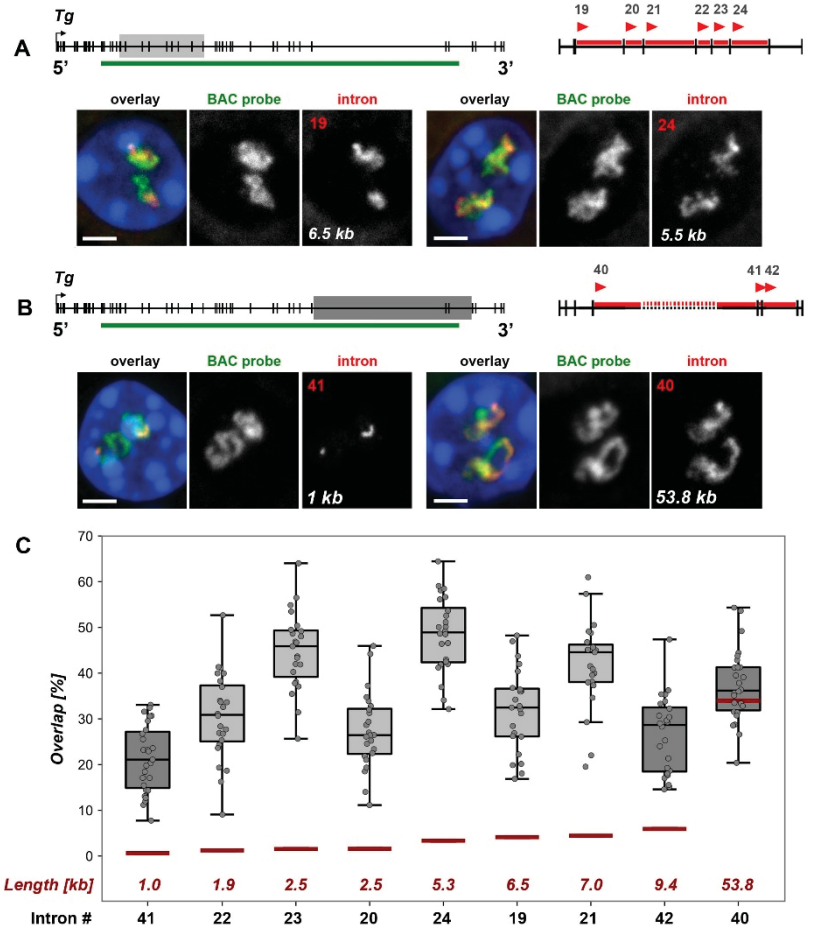


Fig. S1. Overlapping between RNA signals of the *Tg* introns and BAC probe encompassing the middle of the gene (supplementary to Fig. 2).

Pairwise comparison of RNA signals of a single intron *versus* BAC probe (153 kb) in the 5' (**A**) and 3' (**B**) gene regions. The regions are marked on gene schematics with *light-grey* and *dark-grey rectangles*, respectively. The *green lines* below gene schematics indicate localization of the BAC probe; the red arrowheads mark 5'-end of introns within the indicated regions, for which oligoprobes were designed.

Images are projections of confocal stacks through thyrocytes after RNA-FISH with pairwise probes (BAC, *green*; introns, *red*). For clarity, RGB images are appended with grey scale images of RNA signals. Numbers and lengths of introns are indicated on the grey scale intron panels. Scale bars: *2 µm*.

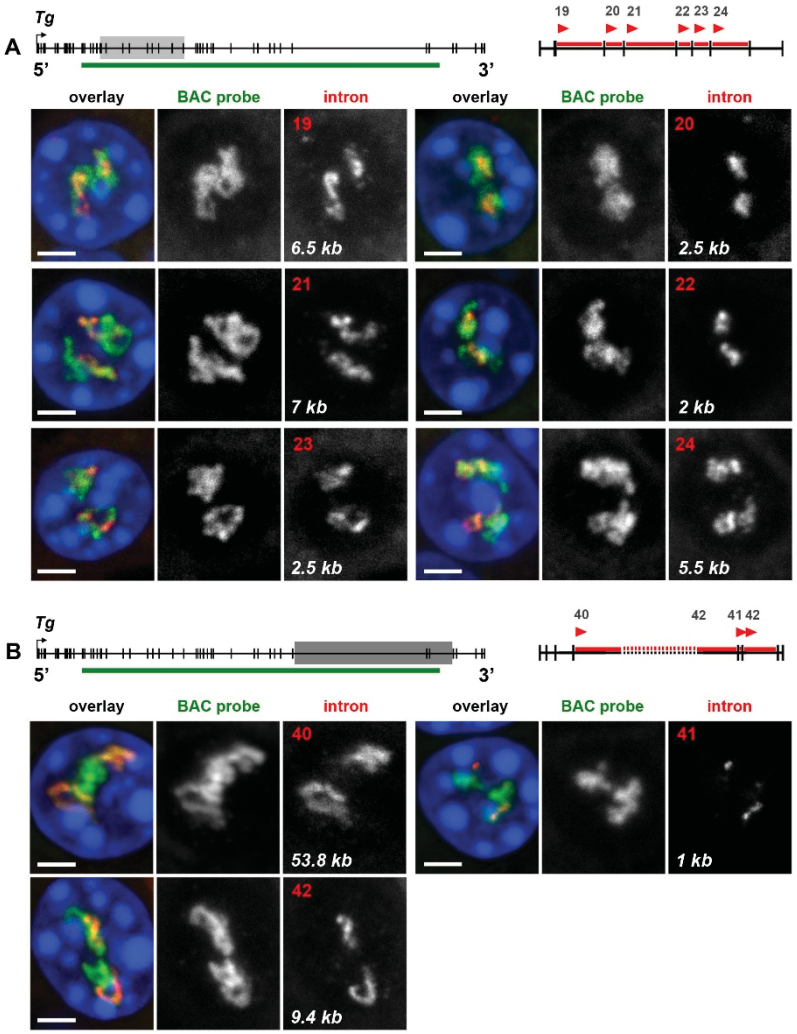


Fig. 3. Schematics of possible splicing scenarios (A-D) and of RNA-FISH experiments for estimation of splicing delay (E).

Strictly co-transcriptional splicing (**A**) implies that introns are excised after polymerases cross the 5' exon border. Singular (**B**) or consecutive (**C**) introns are excised synchronously on a certain distance from a transcription site. Introns are excised in random nRNAs (**D**) in a non-synchronous manner. Note that (B), (C) and (D) scenarios are not mutually exclusive.

Delay in splicing can be estimated by the colocalization degree of RNA-FISH signals (**E**) through comparison of a *reference*-intron (*green*) with *test*-introns (*red*). High signal colocalization (*yellow spot*) means delay in intron excision; absence of signal colocalization (separated *green* and *red spots*) indicates that the *reference*-intron is already excised from that *test*-intron.

ex, exons; *int*, introns; *asterisks* mark spliced transcripts. Schematics of nRNAs (*vertical lines*) are not to scale to the gene regions (*horizontal lines*).

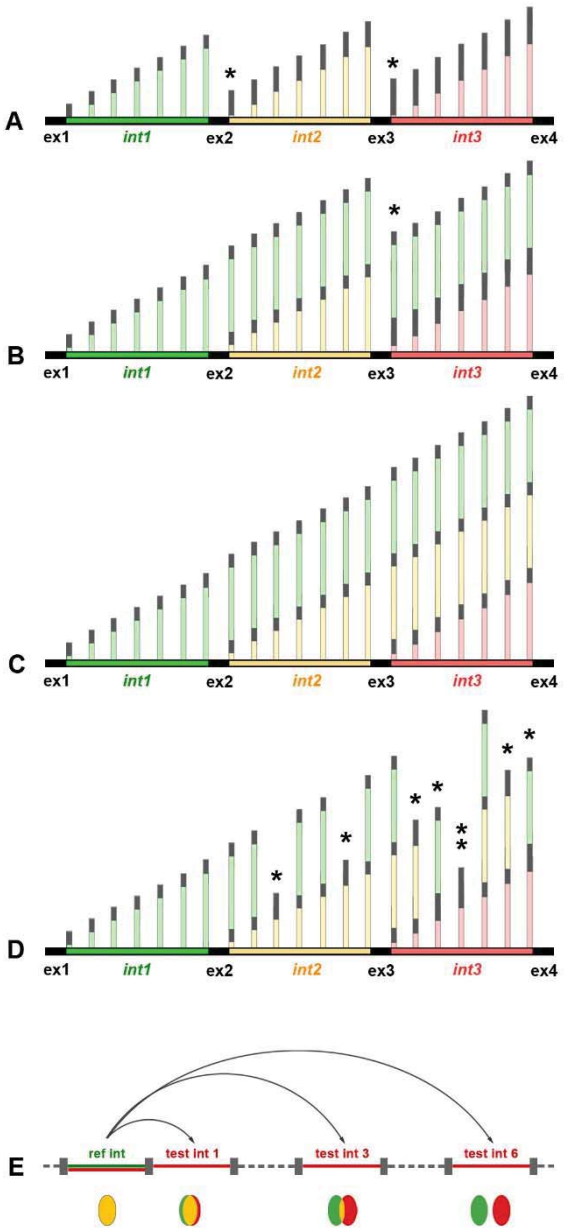


Fig. S2. *Tg* introns remain unspliced for at least 0.75 kb.

A, Gene schematics showing localization of the tested *intron-exon-intron* sequences. In the enlarged view, the used introns are marked with red lines.

B, PCR with primers designed from both sides of the tested exons: ca. 250 bp upstream of an exon 5' border and ca. 750 bp downstream of its 3' border. For all 4 tested exons, the *intron-exon-intron* sequences were successfully amplified.

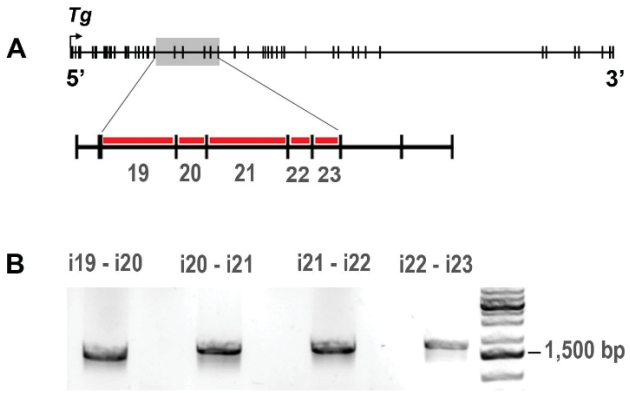
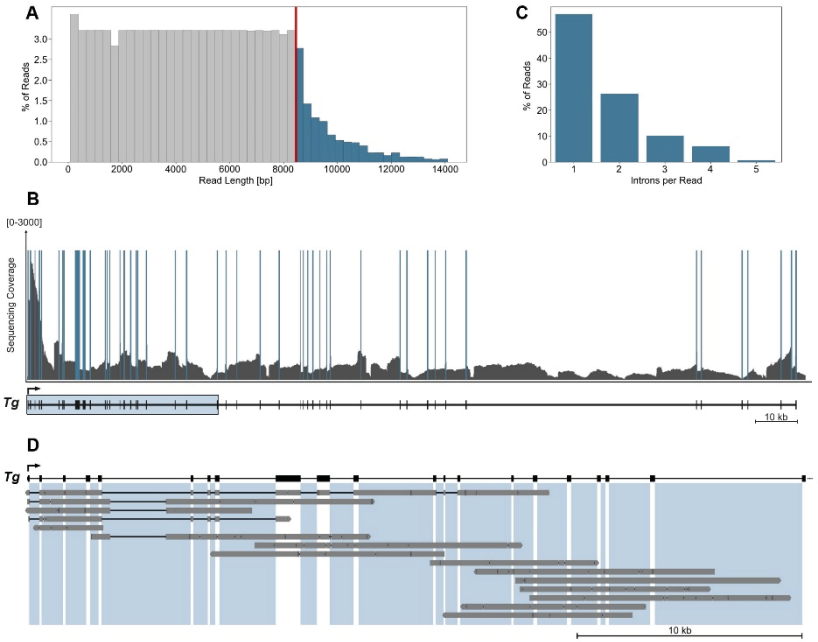


Fig.4. Nanopore sequencing confirms massive retardation of *Tg* introns.



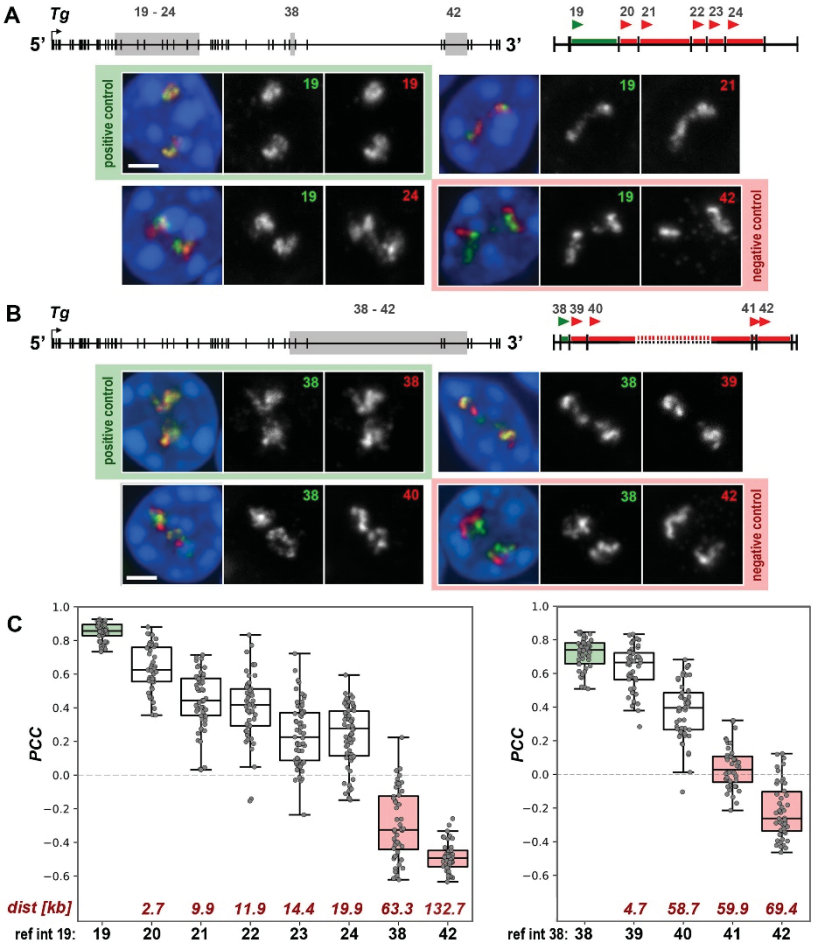
A, length distribution of nascent *Tg* RNA sequencing reads containing introns. X-axis represents the length of the individual sequencing reads. Each bar in the histogram corresponds to the frequency of reads with length intervals of 300 bp. Y-axis represents the proportion of reads. The red line designates the length of a fully spliced *Tg* mRNA (8462 bp).

B, sequencing coverage of nascent RNAs along the Tg gene. *Solid blue bars* depict exons, *grey lines* depict introns. The *Tg* gene schematic is shown below the graph. The Y-axis is cut-off at 3000 reads to highlight the intron coverage.

C, distribution of intron calls per intron containing read. X-axis represents the number of introns contained within individual sequencing reads. Each bar on the histogram represents the frequency of reads. Several reads with 6-8 introns are not included in the graph.

D, graph exemplifying several individual read alignments mapped to the 5' region of the Tg gene, outlined by the box in B. The *thick grey lines* represent exons and unspliced introns; the *thin black lines* represent spliced out regions. *Areas shaded with light-blue* highlight intronic regions within the gene sequence.

Fig. 5. Pairwise comparison of *Tg* intron signals confirms splicing delay for up to 50 kb.



A, B, pairwise comparison of RNA signals between single introns. Regions of intron sampling are marked on gene schematics by *grey rectangles*. On the close-ups of the sampling regions, *green lines* and *red lines* mark the reference- and test-introns; *arrowheads* mark 5'-end of introns, for which oligoprobes were designed.

Exemplifying images of thyrocytes after RNA-FISH with pairwise intron probes are shown below the schematics. The reference-introns 19 (A) and 38 (B) are *green*; the test-introns are *red*. For clarity, the overlaid RGB images of RNA signals are supplemented with grey scale images. Numbers of introns are indicated on the grey scale panels. Images are projections of short confocal stacks. Scale bars: *2 µm*. For more examples of nuclei in this experiment, see Fig. S3.

C, boxplots showing the PCC values for pairwise intron comparison for regions on A (*left*) and B (*right*). Introns are arranged along the X-axis according to their genomic positions. Numbers below the axis are intron numbers; numbers above axis are distances of test-introns from the reference-intron. Thick lines within boxes indicate the medians; boxes show 25th and 75th percentiles; whiskers represent the smallest and the largest values within 1.5 times the interquartile range; grey dots are individual data points. Number of evaluated nuclei ranged between 40 and 50.

Positive and negative controls in are highlighted with *light-green* and *light-red* colors, respectively.

Fig. S3. Pairwise comparison of *Tg* intron signals (supplementary to Fig. 5).

A, B, exemplifying images of thyrocytes after RNA-FISH with pairwise comparison of RNA signals between single introns. Two regions with detected introns at the 5’ end (A) and 3’ end (B) are marked on gene schematics by *grey rectangles*. On the close-ups of the sampling regions, *green lines* and *red lines* mark the reference-introns and test-introns, respectively; arrowheads mark 5’-end of introns, for which oligoprobes were designed.

For clarity, RGB images of overlaid RNA signals are appended with grey scale images. Numbers of introns are indicated on the grey scale panels. Positive and negative controls in A-C are highlighted with *light-green* and *light-red* colors, respectively. Images are projections of short confocal stacks. Scale bars: *2 µm*.

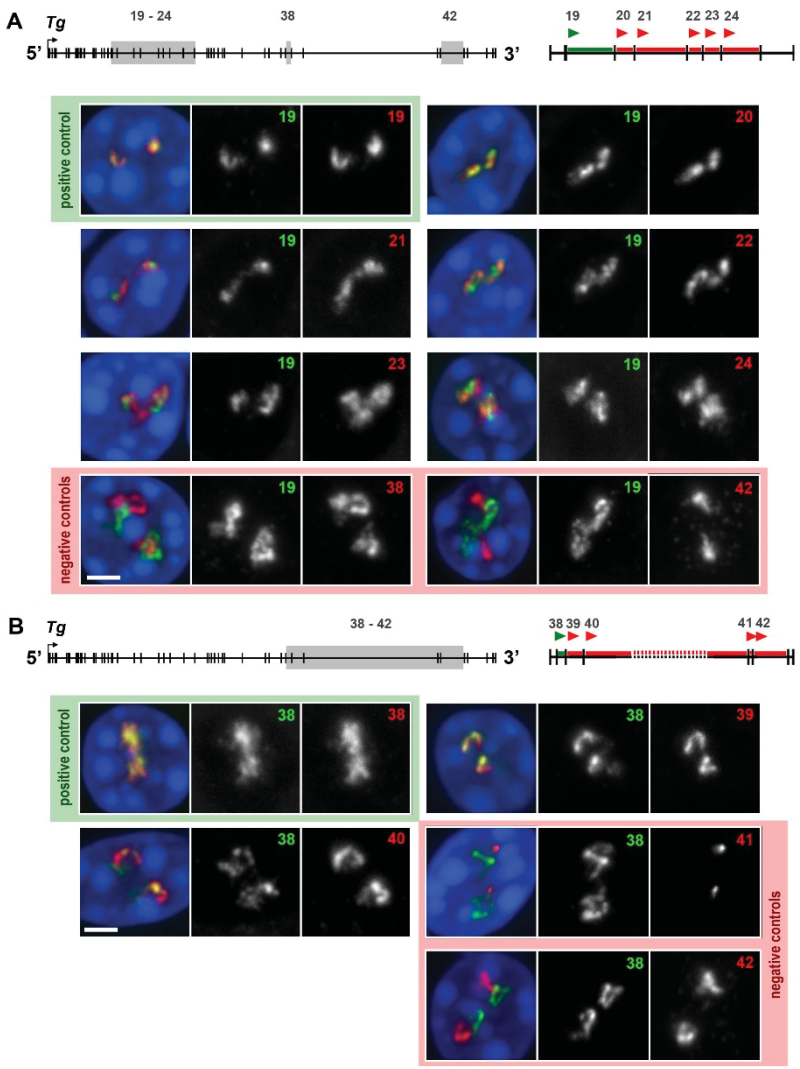
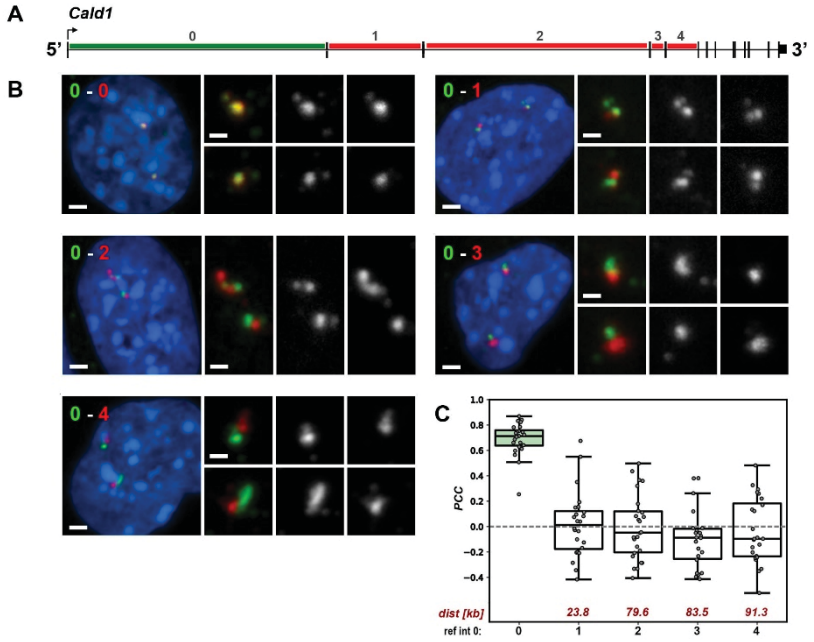


Fig. 6. Pairwise comparison of Cald1 intron signals confirms lack of splicing delay for moderately expressed gene.

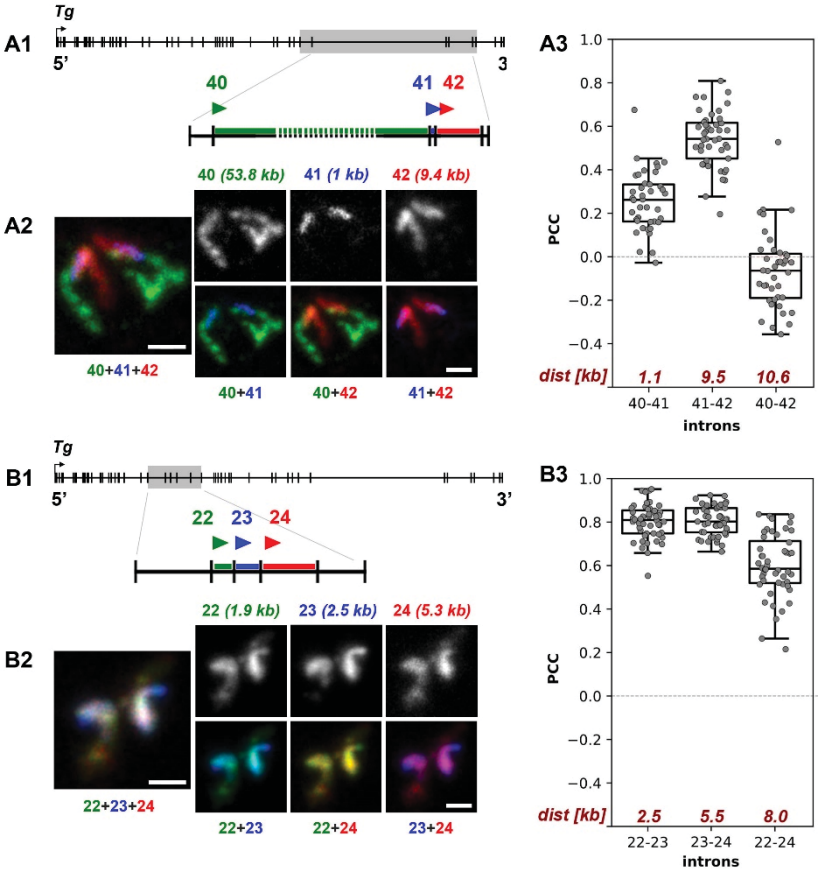


A, schematics of the Cald1 gene with marked reference intron (*green line)* and test-introns (*red lines*). Intron numbers are indicated above the gene axis.

B, exemplifying images of myoblasts after RNA-FISH with pairwise intron probes. The reference-intron 0 is *green,* the test-introns 1-4 are *red*. Numbers of compared introns are indicated in the upper left corners. Close-ups of the signals are shown on the right as RGB overlays and grey scale images. Images are projections of short confocal stacks. Scale bars in panels with nuclei: *2 µm*, in close-ups: *1 µm*.

C, boxplots showing the PCC values for pairwise intron comparison with introns arranged along the X-axis according to their genomic positions. Numbers below the axis are intron numbers; numbers above the axis are genomic distances of test-introns from the reference-intron. Thick lines within boxes indicate the medians; boxes show 25th and 75th percentiles; whiskers represent the smallest and the largest values within 1.5 times the interquartile range; grey dots are individual data points. Number of evaluated nuclei ranged between 22 and 26.

Fig. 7. Long introns exhibit only small delay in splicing.



Estimation of colocalization between three consecutive introns that include one long intron (*53.8 kb*) followed by two short ones (*1 and 9.4 kb*) (**A**) and between three consecutive short introns (*1.9, 2.5 and 5.3 kb*) (**B**).

A1 and B1 are schematics of the gene with highlighted regions of the tested introns also showing the color code. A2 and B2 represent telling examples of intron colocalization. In addition to RGB pairwise overlays and an overlay of all three channels, grey-scale images of single introns are shown. Images are projections. Scale bars: *1 µm*.

Boxplots in A3 and B3 show the PCC values for pairwise intron comparison. Numbers below the X-axis show intron numbers of compared introns; numbers above X-axis indicate distances between them. Thick lines within boxes indicate the medians; boxes show 25th and 75th percentiles; whiskers represent the smallest and the largest values within 1.5 times the interquartile range; grey dots are individual data points. Number of evaluated nuclei ranged between 40 and 50.

Note, that by both visual inspection of the RNA-FISH signals and PCC measurement, the longest *Tg* intron 40 shows little correlation with the following small intron 41 and no correlation with intron 42, whereas both short introns 41 and 42 colocalize strongly (A), similarly to the three short introns in another gene region (B).

3 Discussion and Future Perspectives

3.1 Transcription loop formation as the universal principle of transcription

Although transcription, the pivotal step in gene expression, has received considerable attention in modern research, the spatial organization of individual transcribed genes has remained elusive for a long time. While transcription studies have revealed molecular mechanisms responsible for the regulation of transcription, like promoter-enhancer interactions (Furlong and Levine, 2018), histone modifications (Peterson and Laniel, 2004), and RNA polymerase II modifications (Zaborowska et al., 2016), the understanding of the three-dimensional structure of single transcribed genes has been surprisingly limited.

However, recent research has shed light on different types of chromatin loops and their diverse roles in genome organization and gene regulation. In mitosis, chromatin is extensively compacted into consecutive, non-overlapping loops, facilitated by condensin's loop-extruding activity (Gibcus et al., 2018). Conversely, during interphase, chromatin is structured into loops by cohesin and CTCF, organizing the genome into functional units and contributing to the formation of chromosomal domains such as TADs (Fudenberg et al., 2016; Mirny et al., 2019). These loops facilitate interactions between distal regulatory elements, such as enhancers and promoters, thereby regulating gene transcription (Li et al., 2012).

Transcription loops, unlike the extruded loops mentioned above, are formed not for transcription but as a result of transcription itself (Leidescher et al., 2022). While similar structures were initially described in the late 19th – beginning of 20th century as loops within large chromosomes like lampbrush chromosomes and polytene chromosomes (Callan, 1986; Daneholt, 1975), such structures received little further attention in modern research until recently. Our work revives the classical cytological concepts of lateral loops and puffs observed in lampbrush and polytene chromosomes, bridging these observations with our modern molecular understanding of transcription. Our study has been pivotal in understanding the transcription loop structure and showing the generality of this eukaryotic transcription mechanism.

The fact that transcription loops were described only recently can be mainly attributed to the resolution limitations of conventional light microscopy, because of which TLs largely escaped visual detection by FISH. This limitation arises because most genes under investigation were either too small or too lowly expressed to yield a resolvable loop structure. However, by focusing on long and highly expressed genes, we were able to overcome this limitation and showed that these genes exhibit pronounced decondensation and form discernible loop-like structures when they are expressed. As RNA polymerases traverse along the gene, they generate nascent RNA molecules that are wrapped

into nRNPs, which in turn impart stiffness to the gene axis. This stiffness causes the genes' expansion, resulting in the formation of loops. Thus, in the case of transcription loops, the transcription process itself is the responsible force for loop generation and not *vice versa*.

We could primarily draw these conclusions from imaging data, which provided direct visual evidence of the loop structures. Through fluorescence microscopy, we were able to observe the physical arrangement of the chromatin, follow the process of transcription along the loop, measure inter-flank distances, and follow TL dynamics after manipulation of transcription. Additionally, chromosome and gene modeling played a complementary role by incorporating the simulation of a genes' structure depending on variable gene parameters, such as stiffness or the decoration with nRNAs. By combining experimental and computational approaches, we gained a comprehensive understanding of the mechanisms underlying transcription loop formation. This model of transcription with moving RNAPII along the template DNA opposes the transcription factory model, which states that genomic DNA is reeled through immobilized RNAPII clusters (Papantonis and Cook, 2013). Our research has shown that transcription loop formation represents a universal principle of transcription, as evidenced by the observation of TLs formed by multiple genes across various vertebrate species and in both tissues and cultured cells. Overall, the (re-)discovery of the concept of transcription loop formation represents a significant advancement in our understanding of gene expression and chromatin organization.

3.2 The thyroglobulin gene as a model for perpetually upregulated, high expression

As previously described, the formation of microscopically resolvable transcription loops requires two conditions: the gene must be both relatively long and highly transcribed (Leidescher et al., 2022). However, the combination of high expression and long gene length is extremely rare, especially in cultured mammalian cells, which constitute our primary source of knowledge about transcription. Consequently, TLs had not been described until our serendipitous discovery of TLs formed by the unusually upregulated thyroglobulin gene.

By investigating this model gene further, we have demonstrated that the thyroglobulin gene serves as a hallmark of functional thyrocytes, exhibiting remarkably high-sustained expression levels in these cells. This persistent *Tg* upregulation stands as a defining characteristic of functional thyrocytes and plays a pivotal role in thyroid hormone production. Our investigation included a comprehensive analysis of thyroglobulin gene expression across diverse mouse models with thyroid disorders, as well as thyroid tissue samples from multiple species and even organoids. The high expression of *Tg* and consequently the formation of TLs appear to be conserved across various vertebrate classes, including mammals, birds, frogs, and fish, despite significant differences in gene length across species.

We were able to show that the *Tg* gene does not exhibit rhythmic expression patterns and maintains high levels of upregulation consistently over a 24-hour cycle. Remarkably, our findings revealed that regardless of the underlying condition or genetic background, the thyroglobulin gene maintained robust high expression levels starting from the early stages of thyrocyte differentiation. Even in cases of severe hyperthyroidism or thyroid dysfunction, the gene remains highly expressed. Importantly, the high expression of *Tg* is exclusively observed within thyrocytes gathered into follicle structure. When individual thyrocytes are isolated from the thyroid gland and cultured separately, they lose their distinctive identity and may undergo dedifferentiation and re-enter the cell cycle as previously documented in thyrocyte cultures from other vertebrates (Kimura et al., 2001).

Our study underscores the importance of investigating biological processes within their natural context to gain meaningful insights into gene regulation. While extremely elevated gene expression levels are often associated with artificial systems resulting from manipulations or genetic alterations in cultured cells, studying tissues offers a more physiologically relevant and intricate microenvironment, which cannot be fully replicated *in-vitro*. Therefore, exploring biological processes within tissues is essential for our understanding of gene regulation.

Furthermore, the remarkable expression pattern of the thyroglobulin gene underscores its unique characteristics as a model for studying sustained high upregulation of transcription within a more natural context than conventional cell culture models can offer. Admittedly, there are still significant gaps in our understanding of the mechanisms that drive its unique expression pattern. While it's known that transient overexpression of the transcription factors Nkx2-1 and Pax8 in mouse embryonic stem cells, followed by 3D culturing, can generate follicles that are similar to native thyroid follicles (Antonica et al., 2012), the specific elements responsible for maintaining of high *Tg* expression levels for the entire cell life remain unknown. Understanding the role of natural (super)-enhancers associated with the gene would be of particular interest. Validating putative enhancer elements becomes challenging due to the limited availability of thyroid data and the absence of cell culture models. As a result, the full regulatory landscape involved in *Tg* expression regulation remains unknown.

In addition, the notion of sustained and robust expression of the *Tg* gene under all conditions, at any age, and throughout the day appears counterintuitive and inefficient, which also poses an intriguing biological question that remains unresolved. On one hand, this phenomenon might just reflect an inefficient way of hormone production that evolved during vertebrate evolution. On the other hand, it might serve as an intelligent storage strategy for the rare trace element iodine (Baudry et al., 1998).

Understanding the drastic and dynamic changes in chromatin structure associated with the perpetual gene upregulation is crucial for unraveling the mechanisms driving sustained gene expression. Future research efforts aiming at elucidating these questions can be based on utilizing the thyroglobulin gene as a robust model with a clear microscopic readout.

3.3 Investigating splicing dynamics by microcopy

Another advantage of studying tissues lies in their heterogeneous composition of multiple cells, which exist in diverse states, allowing for the investigation of dynamic processes within the population, even when the tissue as a whole is in a fixed state. The introduction of oligoprobes and the application of techniques like SABER-FISH for signal amplification (Kishi et al., 2019) represent a significant advancement in our ability to study finer details of gene expression. In particular, the remarkable size of *Tg* TLs makes them a unique and compelling model for investigating transcription and splicing processes using light microscopy techniques.

For instance, by utilizing RNA-FISH with oligoprobes covering an intron positioned at the 3' end of *Tg*, we were able to show the presence of transcriptional bursting within the fixed sample. Two equal RNA-FISH signals indicate two actively transcribing alleles; the presence of differently sized signals indicates instances of asynchronous transcriptional bursting. Conversely, in cases where RNA signals were absent, that specific intron was in a transcriptional pause during tissue fixation. Splicing is another crucial aspect of gene expression regulation. In yeast, splicing happens almost immediately upon the synthesis of introns (Oesterreich et al., 2016), whereas in mammals, the exact timing remains a topic of debate (Drexler et al., 2020; Nojima et al., 2018). Our LM and EM studies indicate that TLs are surrounded by nuclear speckles, pointing to a high level of ongoing co-transcriptional splicing activity within these regions. At the same time, RNA-FISH signal from most investigated individual *Tg* introns appeared disproportionately large when compared to the signal of the whole gene, which raised questions about the possibility of a splicing delay occurring during massive transcription.

An ideal method to confirm a delay in splicing would be sequencing-based, however this approach comes with its own set of challenges. Bulk RNA-seq data produced by the most commonly used method of Illumina RNA sequencing poses challenges primarily due to the intrinsically short RNA-seq reads. These short reads are insufficient for characterizing even a single, let alone introns within an individual transcript. Gold-standard techniques for studying ongoing transcription, such as GRO-seq (Core et al., 2008), NET-seq (Churchman and Weissman, 2012), and TT-Seq (Schwalb et al., 2016), all require metabolic labeling or the isolation of chromatin-associated RNA. However, these requirements present significant hindrances for cells in tissues, especially when the tissue is inherently difficult to dissociate and cultivate primary cells while preserving their original transcriptional program, which is particularly the case for the thyroid gland.

Nanopore sequencing theoretically allows for long-read direct RNA sequencing of nascent transcripts and has already been proven as a powerful tool for dissecting the intricacies of nascent RNA processing dynamics (Garalde et al., 2018). In the case of thyroid tissue, drawing reliable conclusions from nanopore sequencing data has proven challenging. Our attempts to directly sequence RNAs extracted from thyroid tissue encountered several challenges. One major issue was the tendency of RNAs to form convoluted secondary and tertiary structures (Zuker and Sankoff, 1984), which often led to clogging of the nanopores and hindered successful sequencing of the native RNA molecules. Additionally, the protocol recommended large quantities of RNA, posing another obstacle to our sequencing efforts caused by the small amount of nascent RNA available in the tissue. Based on rough estimations the necessary quantity of thyroids, and thus the amount of sacrificed mice, to satisfy the minimum required amount of nRNA recommended by the manufacturer for direct RNA sequencing exceed over 200 thyroids and as such surpassed a reasonable number of mice for such an experiment.

These drawbacks lead us to adopt an unconventional nanopore sequencing workflow by creating a sequencing cDNA library which is enriched for nRNAs through poly(A)-depletion of the total RNA fraction isolated from the thyroid tissue. The maximum read length attainable using this method depends on the processivity of the reverse transcriptase employed during the RNA-to-cDNA conversion. However, even with the most advanced reverse transcriptases available today, the fragment length is typically limited to 10-20 kb at most. Still, our analysis of individual reads revealed cases where delayed splicing was clearly evident in individual transcripts, as multiple unspliced introns appeared in a single read, sometimes alongside already spliced introns.

Therefore, we sought another approach to estimate the magnitude of the splicing delay over longer distances and consequently attempted to assess the delay by utilizing RNA-FISH with differentially labeled nRNAs of a reference intron and test introns which are located at different genomic distances from the reference intron. The introns were detected in two spectrally separated channels, enabling pixel intensity-based co-localization analysis. This analysis hinges on the observation that introns are only detectable when they are part of an accumulation of nascent RNAs at the site of transcription, once the intron is spliced out, the signal of the individual intron diminishes or vanishes completely. If splicing occurs immediately, no correlation between a test and subsequent introns should be detectable. However, if splicing is delayed, the signals will correlate until the eventual removal of the test intron from the nascent transcript.

Using this single-cell microscopy method, we were able to provide an estimation of the splicing delay during *Tg* transcription for up to 50- 60 kb. Intriguingly, the longest intron of *Tg*, spanning more than 50 kb also spliced without a delay, indicating that splicing dynamics might depend on the

intron length. The observed delay in splicing of short *Tg* introns also contrasts with the absence of similar delays in the processing of other genes from our thyrocyte nanopore sequencing data. One plausible explanation for this phenomenon would be that that the high transcriptional activity of the *Tg* gene leads to a very localized deficiency of the splicing machinery, without affecting the processing of nRNAs in other genes. Indeed, introns of the moderately expressed *Cald1* gene are spliced promptly, without a delay.

Identifying the exact cause of this delay presents a challenge due to the convoluted nature of the multi-megadalton ribonucleoprotein complex assembling from five U-rich small nuclear RNAs and around 200 associated proteins. The composition and conformation are highly dynamic as the spliceosome has to transition through sequential assembly stages during its functional process. It's plausible that one or multiple factors involved in this process could serve as the rate-limiting step, leading to the delayed splicing observed during the extensive transcription of the *Tg* gene. This theory also aligns with our earlier observation of the massive expansion of the longest *Tg* intron, which is most likely caused by the large and bulky nRNPs that are formed during the transcription of this lengthy intron (Leidescher et al., 2022). This massive expansion essentially spatially segregates the intron from the rest of the *Tg* gene. This spatial separation brings the part of the gene into a microenvironment where the splicing machinery is less exhausted, potentially allowing for more efficient splicing. Additionally, considering that transcription of the entire 53.8 kb intron, based on estimates of a transcription speed of 3.8 kb/min (Singh and Padgett, 2009), must take more than 14 minutes, the splicing process may have sufficient time to "catch-up" with the massive transcription, leading to the timely splicing of the large intron compared to shorter introns. Given the lack of thyrocyte cell culture models, validating all these possible responsible elements and scenarios for the delayed splicing during the expression of *Tg* would necessitate extensive knock-in or knock-out experiments in live animals, which are beyond the scope of this thesis and also not a particular focus of our work.

More importantly, our novel approach to study splicing dynamics underscores the advantage of utilizing TLs as a valuable model for studying transcription and other related processes. By incorporating microscopy observations at the single-cell level, our approach expands the investigative toolbox beyond the already established, mostly sequencing-based methods. This multidimensional approach offers a more comprehensive understanding of the dynamic interplay between transcription, splicing, and chromatin organization within the cellular context.

3.4 The ultrastructure of nascent ribonucleoprotein particles

As described, the bulkiness of nascent ribonucleoprotein particles appears to be the underlying mechanism driving the formation of transcription loops. Unfortunately, our understanding of the

structure of nRNPs remains remarkably limited. Previous studies on RNPs associated with known gene and known mRNA length have been confined to investigations of Balbiani rings, the polytene chromosome puff found in salivary glands of *Chironomus* (Olins et al., 1992). Notably, the mature or nearly mature RNP of the BR2 gene in these rings contains approximately 40 kb of RNA and is densely packed within granules measuring around 50 nm in diameter. In mammals, such studies were conducted on either bulk RNA, or using artificial genes.

It is known that nascent RNA transcripts are enveloped by a complex network of RNA binding proteins (RBPs), which together form the compact nRNPs (Dreyfuss, 1986). Among these RBPs are RNA recognition motif (RRM)-containing SR-proteins, such as the splicing factor SC35, which associates with pre-mRNAs to facilitate splicing (Shepard and Hertel, 2009). Additionally, the exon junction complex (EJC) forms at sites where exons are joined during the splicing process, serving as a binding platform for peripheral factors (Hir et al., 2016). hnRNPC, another prominent RBP, binds to introns within nascent transcripts and significantly influences pre-mRNA compaction. It has been proposed that hnRNPC tetramers act as "RNA nucleosomes," with 150–250 nucleotides of RNA wrapped around them, analogous to DNA wrapped around histones (Huang et al., 1994). Furthermore, the transcription export complex, which links transcription and mRNA export, is co-transcriptionally recruited to nascent transcripts in a splicing-dependent manner in mammals (Reed and Cheng, 2005).

Our research focuses on the overall ultrastructure of nRNPs rather than delving into the specific content of these particles. For this, we utilize the *Ttn* gene (Figure 4A), which stands out in the mammalian genome due to its unusually high ratio of exon to intron lengths, reaching 0.6, and is translated to the biggest mammalian protein titin (*TTN*, TTN). The *Ttn* nRNA steadily extends to a length of 101 kb, making it an exceptional model for studying the ultrastructure of nRNPs, as well as mature mRNAs and their export from the nucleus.

Preliminary electron microscopy observations of myotube nuclei, expressing *Ttn,* have revealed extensive accumulations of granules in the nuclear interior, presumably representing sites of ongoing *Ttn* transcription. The size of most of the granules is around 30 nm in diameter, with some measuring up to 50 nm (Figure 4B). Our hypothesis is supported by several key observations. Firstly, undifferentiated myoblasts lacking *Ttn* expression do not display these granular accumulations. Secondly, when transcription is inhibited in myotubes using chemical inhibitors, there is a notable decrease in both the number of nuclei containing granules and the size of the accumulations. Lastly, the consistent localization of the granule accumulations within nuclear speckles further supports this hypothesis.

Moreover, smaller and more compact granular accumulations are found not in the nuclear interior, but closer to the nuclear envelope, presumably representing single *Ttn* mRNAs ready for

export from the nucleus (Figure 4C). Although direct evidence is currently unavailable due to several technical difficulties, we are actively developing protocols for correlative microscopy with the aim to directly associate the observed granular accumulations with the *Ttn* gene transcription loops (Figure 4A). In collaboration with Dušan Cmarko and his colleagues at Charles University Prague, we are pursuing the identification of granular accumulations through techniques such as in situ hybridization (ISH) on ultrathin sections or focused ion beam (FIB) microscopy. Our objective is to reveal the ultrastructure of nascent ribonucleoprotein particles (nRNPs), aiming to gather deeper insights into the mechanism and architecture of transcription loop organization.

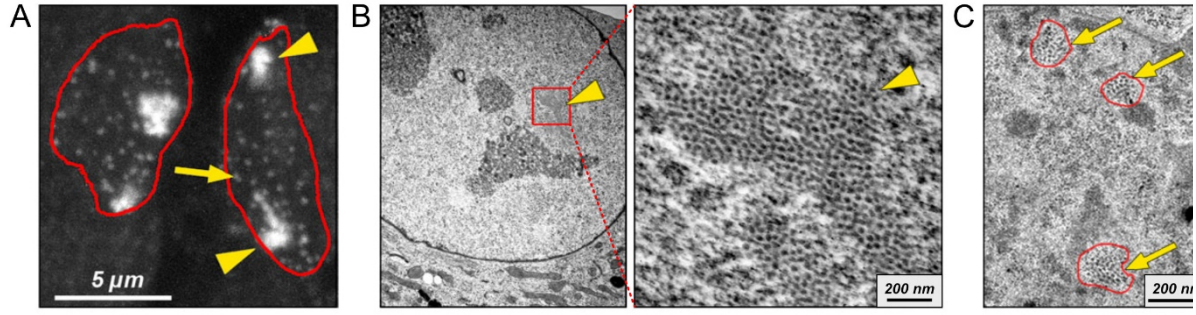


Figure 4: Transcription loop structures and granular accumulations within myotube nuclei.

(A) *Ttn* transcription loop structures (yellow arrow heads) and individual *Ttn* mRNAs (yellow arrows) visualized using RNA FISH. **(B)** Large accumulations of granules in the nuclear interior visualized by electron microscopy. These accumulations are speculated to represent the site of ongoing transcription of the *Ttn* gene. **(C)** Compact granular accumulations located near the nuclear envelope, which likely represent individual *Ttn* mRNAs poised for export from the nucleus.

3.5 The dynamics of transcription loops

As transcription loop formation, and the loop structure itself are likely dynamic processes. A direct visualization of TLs in living cells would offer a valuable approach to study their dynamic nature. Live imaging of RNAs often relies on viral-derived systems such as MS2/MCP or PP7/PCP, which supposedly offer a robust and established method (Tutucci et al., 2018; Wu et al., 2012). These systems entail incorporating stem loop structures into the RNA sequence, which can then be detected using fluorescently labeled binders. While these systems are considered as the gold standard tools for visualizing the dynamics and distribution of mature mRNAs at the single-molecule level, they are not widely employed for visualizing nascent RNA of an entire transcribed gene. Our goal was to adapt this technology to enable the visualization of nascent RNAs along the entire transcription loop of an allele in cultured cells.

To address this, we positioned 24 stem loop repeats near the 5’ end of the target genes. We utilized CRISPR/Cas9 gene editing to insert the array of MS2 loops into the first exon of the *Cald1* gene, a gene expressed in cultured myoblasts (Figure 5, left). Similarly, for the *Ttn* gene expressed in differentiated myotubes, we inserted the same cassette of 24xMS repeats into the third exon of the

gene (Figure 5, right). These insertion sites of the MS2 arrays were strategically selected based on gene regions that are located close to the transcription start site and offer a high gRNA coverage to ensure more efficient insertion. We specifically selected clones with an insertion into only a single allele to make sure that cells produce vital proteins, in case the bulky 24xMS2 cassette at the 5' end of the RNA obstructs ribosome progression and blocks translation. This strategy ensures that functional wild-type RNAs are still available.

Figure 5: Insertion of 24xMS2 Arrays into the *Cald1* and *Ttn* genes.

Illustration of the first 10kb of the genomic locations where the MS2 loop arrays are integrated. The highlighted red regions indicate the insertion points of the 24x MS2 arrays within the respective first and third exons of the *Cald1* and *Ttn* genes.

Both insertions subsequently underwent thorough validation through PCR screenings and locus sequencing. However, the visualization of the inserted MS2 arrays confronted us with several challenges such as optimizing the expression levels of the MCP, selection of suitable fluorophores and maturation times of the fluorescent signals. Achieving the right balance among these parameters still requires extensive experimentation. The signal to noise ratio can pose a significant challenge in these experiments. High background of unbound fluorescently labeled MCP drowns out the actual signal (Figure 6A, top right). If the MCP levels are too low or if the MCP is not properly imported into the nucleus, only mature mRNAs are visible in the cytoplasm leaving no visible nascent RNAs inside the nucleus (Figure 6A bottom left).

Consequently, we generated multiple poly- and monoclonal cell lines using both random integration and a PiggyBac transposase-based system (Yant et al., 2000) for genomic integration. This process involved manual clone picking and FACS sorting to obtain (single-) cell populations. We utilized various promoters to fine-tune the expression levels of the MCP. Additionally, tandem dimer fusions of the MCP were employed in an attempt to enhance signal specificity. Alongside, we tested several different fluorophores with varying intensity and fluorescence maturation times. This aspect of my project consumed a significant portion of my PhD work time. Despite occasional success in visualizing *Cald1* transcription loops (TLs) using the MS2/MCP system (Figure 6B), the reliability and repeatability of this method have not been consistently achieved despite all optimization efforts. Especially the differentiation of myoblasts into myotubes after knocking in a 24xMS2 array is particularly challenging.

The insertion of the large DNA sequence can cause genomic instability and transcriptional interference, disrupting the normal expression of the *Ttn* gene which crucial for the differentiation process. Additionally, the presence of exogenous sequences like the MS2 array can lead to epigenetic changes, altering chromatin structure and affecting the accessibility of differentiation-related transcription factors. The genetic modifications may also induce cellular stress, hindering the morphological changes necessary for myotube formation. As such this project was partially successful, but remains ongoing to be fully completed in the future.

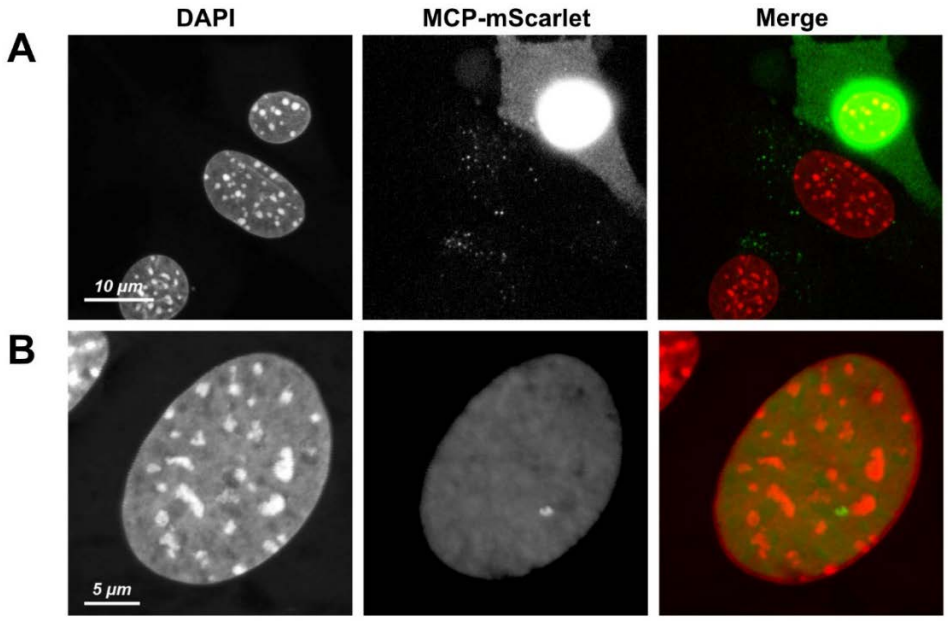


Figure 6: Challenges in visualizing *Cald1* transcription loops in myoblasts using MS2/MCP.

(A) Top right cell: A high amount of fluorescence from unbound MCP can overwhelm the actual signal, hindering the detection of nascent RNAs within the nucleus. Bottom left cells: Conversely, low MCP abundance may lead to the absence of visible nascent RNAs within the nucleus, with only mature mRNAs observable in the cytoplasm. **(B)** Successful visualization of a *Cald1* transcription loop using the MS2/MCP system, as evidenced by the discernible TL structure visible for the single MS2 tagged allele.

3.6 Alternative model systems to consider for the further investigation of transcription loops

Another critical factor to consider for studying TLs is the choice of a model organism. Genomes of commonly used model organisms – such as mouse, human, *C.elegans* or *Drosophila* – are lacking multiple long and highly expressed genes, posing a major limitation in identifying suitable candidate genes for TL studying. This shortage of suitable candidate genes emphasizes a significant challenge in leveraging existing model systems to investigate regulation of highly expressed genes, TL formation

and interaction of multiple TLs. With its 32-gigabasepair genome, the tailed amphibian axolotl (*Ambystoma mexicanum*) represents a promising model to study these transcription features.

The genome of the axolotl contains numerous long and highly expressed genes with long introns, with a median intron size of 23kb, which is approximately 13 times larger than that in humans (Nowoshilow et al., 2018). This characteristic would make such a genome an ideal candidate for investigating several individual or even multiple huge TLs within interior of a single nucleus. Towards this goal, our colleagues from Elly Tanaka's lab (Institute of Molecular Pathology, Vienna) assisted us in selecting genes expressed in various axolotl tissues. We fixed several axolotl tissues - liver, brain, and heart - and made efforts to design oligo-probes for RNA-FISH, targeting genes expressed in these tissues. However, due to the poor annotation of the axolotl's genome and transcriptome, as well as the abundance of repetitive sequences enormously complicating bioinformatical computation of this genome, the visualization of these genes remains to be fully completed.

As such, despite its potential, the axolotl is not yet our model of choice for studying TL formation. Therefore, we have opted for mouse cells and tissues with a better-characterized genome and transcriptome in our studies. However, as more groundwork is laid in genome and transcriptome annotation, the axolotl hopefully will regain its appeal as a model organism for studying TLs in the future. With improved understanding and resources, the unique biological features of the axolotl could make it an indispensable tool for unraveling the intricacies of TL formation and gene regulation.

4 Appendix

4.1 References

Allan, J., P.G. Hartman, C. Crane-Robinson, and F.X. Aviles. 1980. The structure of histone H1 and its location in chromatin. *Nature*. 288:675-679.

Amara, S.G., V. Jonas, M.G. Rosenfeld, E.S. Ong, and R.M. Evans. 1982. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature*. 298:240-244.

Antonica, F., D.F. Kasprzyk, R. Opitz, M. Iacovino, X.-H. Liao, A.M. Dumitrescu, S. Refetoff, K. Peremans, M. Manto, and M. Kyba. 2012. Generation of functional thyroid from embryonic stem cells. *Nature*. 491:66-71.

Baldi, S., P. Korber, and P.B. Becker. 2020. Beads on a string-nucleosome array arrangements and folding of the chromatin fiber. *Nat Struct Mol Biol*. 27:109-118.

Bannister, A.J., P. Zegerman, J.F. Partridge, E.A. Miska, J.O. Thomas, R.C. Allshire, and T. Kouzarides. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*. 410:120-124.

Baudry, N., P.-J. Lejeune, F. Delom, L. Vinet, P. Carayon, and B. Mallet. 1998. Role of multimerized porcine thyroglobulin in iodine storage. *Biochemical and biophysical research communications*. 242:292-296.

Bauman, J., J. Wiegant, P. Borst, and P. Van Duijn. 1980. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Experimental cell research*. 128:485-490.

Beliveau, B.J., A.N. Boettiger, M.S. Avendaño, R. Jungmann, R.B. McCole, E.F. Joyce, C. Kim-Kiselak, F. Bantignies, C.Y. Fonseka, and J. Erceg. 2015. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nature communications*. 6:7147.

Beliveau, B.J., J.Y. Kishi, G. Nir, H.M. Sasaki, S.K. Saka, S.C. Nguyen, C.-t. Wu, and P. Yin. 2018. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proceedings of the National Academy of Sciences*. 115:E2183-E2192.

Bickmore, W.A. 2013. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet*. 14:67-84.

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

Björkroth, B., C. Ericsson, M. Lamb, and B. Daneholt. 1988. Structure of the chromatin axis during transcription. *Chromosoma*. 96:333-340.

Boehning, M., C. Dugast-Darzacq, M. Rankovic, A.S. Hansen, T. Yu, H. Marie-Nelly, D.T. McSwiggen, G. Kokic, G.M. Dailey, P. Cramer, X. Darzacq, and M. Zweckstetter. 2018. RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol*. 25:833-840.

Boettiger, A., and S. Murphy. 2020. Advances in chromatin imaging at kilobase-scale resolution. *Trends in Genetics*. 36:273-287.

Borsari, B., P. Villegas-Mirón, S. Pérez-Lluch, I. Turpin, H. Laayouni, A. Segarra-Casas, J. Bertranpetit, R. Guigó, and S. Acosta. 2021. Enhancers with tissue-specific activity are enriched in intronic regions. *Genome research*. 31:1325-1336.

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

Brero, A., H.P. Easwaran, D. Nowak, I. Grunewald, T. Cremer, H. Leonhardt, and M.C. Cardoso. 2005. Methyl CpG–binding proteins induce large-scale chromatin reorganization during terminal differentiation. *The Journal of cell biology*. 169:733-743.

Brown, D.D. 1984. The role of stable complexes that repress and activate eucaryotic genes. *Cell*. 37:359-365.

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

Bulger, M., and M. Groudine. 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell*. 144:327-339.

Cai, D., Z. Liu, and J. Lippincott-Schwartz. 2021. Biomolecular condensates and their links to cancer progression. *Trends in biochemical sciences*. 46:535-549.

Callan, H.G. 1986. Lampbrush chromosomes. *Mol Biol Biochem Biophys*. 36:1-252.

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

Churchman, L.S., and J.S. Weissman. 2012. Native elongating transcript sequencing (NET-seq). *Current protocols in molecular biology*. 98:14.14. 11-14.14. 17.

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

Core, L.J., J.J. Waterfall, and J.T. Lis. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 322:1845-1848.

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

Cremer, T., and C. Cremer. 2006. Rise, fall and resurrection of chromosome territories: a historical perspective Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and m. *European journal of histochemistry*. 50:223-272.

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

Daneholt, B. 1975. Transcription in polytene chromosomes. *Cell*. 4:1-9.

De Laat, W., and D. Duboule. 2013. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*. 502:499-506.

de Wit, E., E.S. Vos, S.J. Holwerda, C. Valdes-Quezada, M.J. Verstegen, H. Teunissen, E. Splinter, P.J. Wijchers, P.H. Krijger, and W. de Laat. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*. 60:676-684.

Dejosez, M., A. Dall'Agnese, M. Ramamoorthy, J. Platt, X. Yin, M. Hogan, R. Brosh, A.S. Weintraub, D. Hnisz, and B.J. Abraham. 2023. Regulatory architecture of housekeeping genes is driven by promoter assemblies. *Cell reports*. 42.

Deng, W., J. Lee, H. Wang, J. Miller, A. Reik, P.D. Gregory, A. Dean, and G.A. Blobel. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*. 149:1233-1244.

Dixon, J.R., I. Jung, S. Selvaraj, Y. Shen, J.E. Antosiewicz-Bourget, A.Y. Lee, Z. Ye, A. Kim, N. Rajagopal, and W. Xie. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 518:331-336.

Dixon, J.R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, and B. Ren. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 485:376-380.

Dossin, F., I. Pinheiro, J.J. Zylicz, J. Roensch, S. Collombet, A. Le Saux, T. Chelmicki, M. Attia, V. Kapoor, Y. Zhan, F. Dingli, D. Loew, T. Mercher, J. Dekker, and E. Heard. 2020. SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*. 578:455–460.

Dowen, J.M., Z.P. Fan, D. Hnisz, G. Ren, B.J. Abraham, L.N. Zhang, A.S. Weintraub, J. Schuijers, T.I. Lee, and K. Zhao. 2014. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*. 159:374-387.

Drexler, H.L., K. Choquet, and L.S. Churchman. 2020. Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Molecular cell*. 77:985-998. e988.

Dreyfuss, G. 1986. Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. *Annual review of cell biology*. 2:459-498.

Elbaum-Garfinkle, S. 2019. Matter over mind: Liquid phase separation and neurodegeneration. *Journal of Biological Chemistry*. 294:7160-7168.

Erdel, F., and K. Rippe. 2018. Formation of chromatin subcompartments by phase separation. *Biophysical journal*. 114:2262-2270.

Flavahan, W.A., Y. Drier, B.B. Liau, S.M. Gillespie, A.S. Venteicher, A.O. Stemmer-Rachamimov, M.L. Suva, and B.E. Bernstein. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 529:110-114.

Friman, E.T., and W.A. Bickmore. 2022. The sight of transcription. *Nature Cell Biology*. 24:284-285.

Fudenberg, G., M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L.A. Mirny. 2016. Formation of chromosomal domains by loop extrusion. *Cell reports*. 15:2038-2049.

Furlong, E.E., and M. Levine. 2018. Developmental enhancers and chromosome topology. *Science*. 361:1341-1345.

Galupa, R., and E. Heard. 2017. Topologically associating domains in chromosome architecture and gene regulatory landscapes during development, disease, and evolution. *In* Cold Spring Harbor symposia on quantitative biology. Vol. 82. Cold Spring Harbor Laboratory Press. 267-278.

Garalde, D.R., E.A. Snell, D. Jachimowicz, B. Sipos, J.H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James, and A. Warland. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nature methods*. 15:201-206.

Gaszner, M., and G. Felsenfeld. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews Genetics*. 7:703-713.

Gelali, E., G. Girelli, M. Matsumoto, E. Wernersson, J. Custodio, A. Mota, M. Schweitzer, K. Ferenc, X. Li, and R. Mirzazadeh. 2019. iFISH is a publically available resource enabling versatile DNA FISH to study genome architecture. *Nature communications*. 10:1636.

Ghamari, A., M.P. van de Corput, S. Thongjuea, W.A. van Cappellen, W. van Ijcken, J. van Haren, E. Soler, D. Eick, B. Lenhard, and F.G. Grosveld. 2013. In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes Dev*. 27:767-777.

Gibcus, J.H., K. Samejima, A. Goloborodko, I. Samejima, N. Naumova, J. Nuebler, M.T. Kanemaki, L. Xie, J.R. Paulson, and W.C. Earnshaw. 2018. A pathway for mitotic chromosome formation. *Science*. 359:eaao6135.

Gilbert, W. 1978. Why genes in pieces? *Nature.* 271:501.

Goetze, S., J. Mateos-Langerak, H.J. Gierman, W. de Leeuw, O. Giromus, M.H. Indemans, J. Koster, V. Ondrej, R. Versteeg, and R. van Driel. 2007. The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol*. 27:4475-4487.

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

Guo, Y., Q. Xu, D. Canzio, J. Shou, J. Li, D.U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M.Q. Zhang, B. Ren, A.R. Krainer, T. Maniatis, and Q. Wu. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*. 162:900-910.

Hall, L.L., K.P. Smith, M. Byron, and J.B. Lawrence. 2006. Molecular anatomy of a speckle. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology: An Official Publication of the American Association of Anatomists*. 288:664-675.

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

Heard, E. 2005. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Current opinion in genetics & development*. 15:482-489.

Heitz, E. 1928. Das Heterochromatin der Moose, Jahrb. Wiss. Bot.

Hershberg, E.A., C.K. Camplisson, J.L. Close, S. Attar, R. Chern, Y. Liu, S. Akilesh, P.R. Nicovich, and B.J. Beliveau. 2021. PaintSHOP enables the interactive design of transcriptome-and genome-scale oligonucleotide FISH experiments. *Nature methods*. 18:937-944.

Hildebrand, E.M., and J. Dekker. 2020. Mechanisms and functions of chromosome compartmentalization. *Trends in biochemical sciences*. 45:385-396.

Hir, H.L., J. Saulière, and Z. Wang. 2016. The exon junction complex as a node of post-transcriptional networks. *Nature reviews Molecular cell biology*. 17:41-54.

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

Hnisz, D., A.S. Weintraub, D.S. Day, A.L. Valton, R.O. Bak, C.H. Li, J. Goldmann, B.R. Lajoie, Z.P. Fan, A.A. Sigova, J. Reddy, D. Borges-Rivera, T.I. Lee, R. Jaenisch, M.H. Porteus, J. Dekker, and R.A. Young. 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 351:1454-1458.

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

Huang, M., J.E. Rech, S.J. Northington, P.F. Flicker, A. Mayeda, A.R. Krainer, and W.M. LeStourgeon. 1994. The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. *Mol Cell Biol*. 14:518-533.

Hyman, A.A., C.A. Weber, and F. Julicher. 2014. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol*. 30:39-58.

Jackson, D.A., A.B. Hassan, R.J. Errington, and P.R. Cook. 1993. Visualization of focal sites of transcription within human nuclei. *EMBO J*. 12:1059-1065.

Karpinska, M.A., and A.M. Oudelaar. 2023. The role of loop extrusion in enhancer-mediated gene activation. *Current Opinion in Genetics & Development*. 79:102022.

Kim, Y., Z. Shi, H. Zhang, I.J. Finkelstein, and H. Yu. 2019. Human cohesin compacts DNA by loop extrusion. *Science*. 366:1345-1349.

Kimura, T., A. Van Keymeulen, J. Golstein, A. Fusco, J.E. Dumont, and P.P. Roger. 2001. Regulation of thyroid cell proliferation by TSH and other factors: a critical evaluation of in vitro models. *Endocrine reviews*. 22:631-656.

Kishi, J.Y., S.W. Lapan, B.J. Beliveau, E.R. West, A. Zhu, H.M. Sasaki, S.K. Saka, Y. Wang, C.L. Cepko, and P. Yin. 2019. SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. *Nature methods*. 16:533-544.

Kishi, J.Y., T.E. Schaus, N. Gopalkrishnan, F. Xuan, and P. Yin. 2018. Programmable autonomous synthesis of single-stranded DNA. *Nature chemistry*. 10:155-164.

Kristjuhan, A., and J.Q. Svejstrup. 2004. Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo. *The EMBO journal*. 23:4243-4252.

Kupper, K., A. Kolbl, D. Biener, S. Dittrich, J. von Hase, T. Thormeyer, H. Fiegler, N.P. Carter, M.R. Speicher, T. Cremer, and M. Cremer. 2007. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma*. 116:285-306.

Langer-Safer, P.R., M. Levine, and D.C. Ward. 1982. Immunological method for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy of Sciences*. 79:4381-4385.

Lee, C.-K., Y. Shibata, B. Rao, B.D. Strahl, and J.D. Lieb. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature genetics*. 36:900-905.

Leemans, C., M.C. van der Zwalm, L. Brueckner, F. Comoglio, T. van Schaik, L. Pagie, J. van Arensbergen, and B. van Steensel. 2019. Promoter-intrinsic and local chromatin features determine gene repression in LADs. *Cell*. 177:852-864. e814.

Leidescher, S., J. Ribisel, S. Ullrich, Y. Feodorova, E. Hildebrand, A. Galitsyna, S. Bultmann, S. Link, K. Thanisch, and C. Mulholland. 2022. Spatial organization of transcribed eukaryotic genes. *Nature cell biology*. 24:327-339.

Lettice, L.A., S.J. Heaney, L.A. Purdie, L. Li, P. De Beer, B.A. Oostra, D. Goode, G. Elgar, R.E. Hill, and E. de Graaff. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics*. 12:1725-1735.

Li, G., X. Ruan, R.K. Auerbach, K.S. Sandhu, M. Zheng, P. Wang, H.M. Poh, Y. Goh, J. Lim, and J. Zhang. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*. 148:84-98.

Lupianez, D.G., K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J.M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S.A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 161:1012-1025.

Macgregor, H.C. 1993. An introduction to animal cytogenetics. Chapman and Hall, London.

Martin, S., and A. Pombo. 2003. Transcription factories: quantitative studies of nanostructures in the mammalian nucleus. *Chromosome Res*. 11:461-470.

Mateo, L.J., S.E. Murphy, A. Hafner, I.S. Cinquini, C.A. Walker, and A.N. Boettiger. 2019. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*. 568:49-54.

Miller, O.L., Jr., and B.R. Beatty. 1969. Visualization of nucleolar genes. *Science*. 164:955-957.

Mirny, L.A., M. Imakaev, and N. Abdennur. 2019. Two major mechanisms of chromosome organization. *Current opinion in cell biology*. 58:142-152.

Nojima, T., K. Rebelo, T. Gomes, A.R. Grosso, N.J. Proudfoot, and M. Carmo-Fonseca. 2018. RNA polymerase II phosphorylated on CTD serine 5 interacts with the spliceosome during co-transcriptional splicing. *Molecular cell*. 72:369-379. e364.

Nowoshilow, S., S. Schloissnig, J.-F. Fei, A. Dahl, A.W. Pang, M. Pippel, S. Winkler, A.R. Hastie, G. Young, and J.G. Roscito. 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature*. 554:50-55.

Oesterreich, F.C., L. Herzel, K. Straube, K. Hujer, J. Howard, and K.M. Neugebauer. 2016. Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell*. 165:372-381.

Olins, A.L., D.E. Olins, and D.P. Bazett-Jones. 1992. Balbiani ring hnRNP substructure visualized by selective staining and electron spectroscopic imaging. *J Cell Biol*. 117:483-491.

Olins, D.E., and A.L. Olins. 2003. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol*. 4:809-814.

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

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

Peric-Hupkes, D., W. Meuleman, L. Pagie, S.W. Bruggeman, I. Solovei, W. Brugman, S. Gräf, P. Flicek, R.M. Kerkhoven, and M. van Lohuizen. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Molecular cell*. 38:603-613.

Peters, A.H., S. Kubicek, K. Mechtler, R.J. O'Sullivan, A.A. Derijck, L. Perez-Burgos, A. Kohlmaier, S. Opravil, M. Tachibana, Y. Shinkai, J.H. Martens, and T. Jenuwein. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell*. 12:1577-1589.

Peterson, C.L., and M.-A. Laniel. 2004. Histones and histone modifications. *Current Biology*. 14:R546-R551.

Pickersgill, H., B. Kalverda, E. De Wit, W. Talhout, M. Fornerod, and B. Van Steensel. 2006. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nature genetics*. 38:1005-1014.

Player, A.N., L.-P. Shen, D. Kenny, V.P. Antao, and J.A. Kolberg. 2001. Single-copy gene detection using branched DNA (bDNA) in situ hybridization. *Journal of Histochemistry & Cytochemistry*. 49:603-611.

Poleshko, A., and R.A. Katz. 2014. Specifying peripheral heterochromatin during nuclear lamina reassembly. *Nucleus*. 5:32-39.

Raj, A., C.S. Peskin, D. Tranchina, D.Y. Vargas, and S. Tyagi. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 4:e309.

Reed, R., and H. Cheng. 2005. TREX, SR proteins and export of mRNA. *Curr Opin Cell Biol*. 17:269-273.

Rosa, A., and R. Everaers. 2008. Structure and dynamics of interphase chromosomes. *PLoS computational biology*. 4:e1000153.

Scheer, U. 1987. Contributions of electron microscopic spreading preparations ("Miller spreads") to the analysis of chromosome structure. *Results Probl Cell Differ*. 14:147-171.

Schermelleh, L., R. Heintzmann, and H. Leonhardt. 2010. A guide to super-resolution fluorescence microscopy. *J Cell Biol*. 190:165-175.

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

Schoenfelder, S., T. Sexton, L. Chakalova, N.F. Cope, A. Horton, S. Andrews, S. Kurukuti, J.A. Mitchell, D. Umlauf, and D.S. Dimitrova. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature genetics*. 42:53-61.

Schotta, G., M. Lachner, K. Sarma, A. Ebert, R. Sengupta, G. Reuter, D. Reinberg, and T. Jenuwein. 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev*. 18:1251-1262.

Schwalb, B., M. Michel, B. Zacher, K. Frühauf, C. Demel, A. Tresch, J. Gagneur, and P. Cramer. 2016. TT-seq maps the human transient transcriptome. *Science*. 352:1225-1228.

Sealey, P.G., P.A. Whittaker, and E.M. Southern. 1985. Removal of repeated sequences from hybridisation probes. *Nucleic acids research*. 13:1905-1922.

See, K., Y. Lan, J. Rhoades, R. Jain, C.L. Smith, and J.A. Epstein. 2019. Lineage-specific reorganization of nuclear peripheral heterochromatin and H3K9me2 domains. *Development*. 146:dev174078.

Shepard, P.J., and K.J. Hertel. 2009. The SR protein family. *Genome Biol*. 10:242.

Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proceedings of the National Academy of Sciences*. 89:8794-8797.

Shopland, L.S., C.V. Johnson, M. Byron, J. McNeil, and J.B. Lawrence. 2003. Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *The Journal of cell biology*. 162:981-990.

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

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

Spector, D.L. 1993. Nuclear organization of pre-mRNA processing. *Curr Opin Cell Biol*. 5:442-447.

Spector, D.L., and A.I. Lamond. 2011. Nuclear speckles. *Cold Spring Harbor perspectives in biology*. 3:a000646.

Steger, D.J., and J.L. Workman. 1996. Remodeling chromatin structures for transcription: what happens to the histones? *Bioessays*. 18:875-884.

Tammer, L., O. Hameiri, I. Keydar, V.R. Roy, A. Ashkenazy-Titelman, N. Custódio, I. Sason, R. Shayevitch, V. Rodríguez-Vaello, J. Rino, G. Lev Maor, Y. Leader, D. Khair, E.L. Aiden, R. Elkon, M. Irimia, R. Sharan, Y. Shav-Tal, M. Carmo-Fonseca, and G. Ast. 2022. Gene architecture directs splicing outcome in separate nuclear spatial regions. *Mol Cell*. 82:1021-1034.e1028.

Tolhuis, B., R.-J. Palstra, E. Splinter, F. Grosveld, and W. De Laat. 2002. Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular cell*. 10:1453-1465.

Trojer, P., and D. Reinberg. 2007. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell*. 28:1-13.

Tutucci, E., M. Vera, J. Biswas, J. Garcia, R. Parker, and R.H. Singer. 2018. An improved MS2 system for accurate reporting of the mRNA life cycle. *Nature methods*. 15:81-89.

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

Voss, T.C., and G.L. Hager. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet*. 15:69-81.

Waddington, C.H. 1942. The Epigenotype. *Endeavour*. 1:18-20.

Wang, E.T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, and C.B. Burge. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 456:470-476.

Wang, H., L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R.S. Jones, and Y. Zhang. 2004. Role of histone H2A ubiquitination in Polycomb silencing. *Nature*. 431:873-878.

Wansink, D.G., W. Schul, I. van der Kraan, B. van Steensel, R. van Driel, and L. de Jong. 1993. Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J Cell Biol*. 122:283-293.

Warburton, D. 1991. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *American journal of human genetics*. 49:995.

Wei, M.-T., Y.-C. Chang, S.F. Shimobayashi, Y. Shin, A.R. Strom, and C.P. Brangwynne. 2020. Nucleated transcriptional condensates amplify gene expression. *Nature cell biology*. 22:1187-1196.

Workman, J.L. 2006. Nucleosome displacement in transcription. *Genes & development*. 20:2009-2017.

Wu, B., J.A. Chao, and R.H. Singer. 2012. Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. *Biophysical journal*. 102:2936-2944.

Yant, S.R., L. Meuse, W. Chiu, Z. Ivics, Z. Izsvak, and M.A. Kay. 2000. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nature genetics*. 25:35-41.

Zaborowska, J., S. Egloff, and S. Murphy. 2016. The pol II CTD: new twists in the tail. *Nature structural & molecular biology*. 23:771-777.

Zuker, M., and D. Sankoff. 1984. RNA secondary structures and their prediction. *Bulletin of mathematical biology*. 46:591-621.

4.2 Abbreviations

µm	Micrometer
3D	Three dimensional
Ac	Acetylation
BAC	Bacterial artificial chromosome
Bp	Base pair
BR	Balbiani ring
Cald1	*Caldesmon 1*
Cas	CRISPR associated
cDNA	Complementary DNA
cHC	Constitutive heterochromatin
CpG	Cytosine nucleotide followed by a guanine nucleotide
CRISPR	Clustered regularly interspaced short palindromic repeats
CT	Chromosome territory
CTCF	CCCTC-binding factor
CTD	C-terminal domain
DAPI	4’,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
EC	Euchromatin
EM	Electron microscope/microscopy
ERAF	Erythroid associated factor
FACS	Fluorescence activated cell sorting
fHC	Facultative heterochromatin
FIB	Focused ion beam
FISH	Fluorescence in situ hybridization
gRNA	Guide RNA
HC	Heterochromatin
HP1	Heterochromatin protein 1
kb	Kilo base (pair)
LAD	Lamina associated domain
LBC	Lampbrush chromosome
MCP	MS2 bacteriophage coat protein
Me	Methylation
mRNA	messenger RNA
MS2	bacteriophage MS2
nm	Nanometer
nRNA	nascent RNA
nRNP	Nascent ribonucleoprotein particle
Oligo	Oligonucleotides
PAX8	Paired box gene 8
PCP	PP7 bacteriophage coat protein
PCR	Polymerase chain reaction
PP7	bacteriophage PP7
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
SABER	Signal amplification by exchange reaction
SC35	Splicing component 35 kDa
SLC4A1	Solute carrier family 4 member 1

TAD	Topologically associating domain
TF	Transcription factor
Tg	Thyroglobulin
TL	Transcription loop
TPM	Transcripts per million
Ttn	Titin
Ub	Ubiquitination

4.3 List of publications

4.3.1 Published Papers

Leidescher, S., Ribisel, J.*, **Ullrich, S**.*, Feodorova, Y., Hildebrand, E., Galitsyna, A., Bultmann, S., Link, S., Thanisch, K., Mulholland, C., Dekker, J., Leonhardt, H., Mirny, L.** & Solovei, I.** (**2022**). Spatial organization of transcribed eukaryotic genes. Nature Cell Biology, 24(3), 327–339. https://doi.org/10.1038/s41556-022-00847-6

Ullrich, S., Leidescher, S., Feodorova, Y., Thanisch, K., Fini, J., Kaspers, B., Weber, F., Markova, B., Führer, D., Romitti, M., Krebs, S., Blum, H., Leonhardt, H., Costagliola, S., Heuer, H. & Solovei, I. (**2023**). The highly and perpetually upregulated thyroglobulin gene is a hallmark of functional thyrocytes. Frontiers in Cell And Developmental Biology, 11. https://doi.org/10.3389/fcell.2023.1265407

4.3.2 Manuscripts in Preparation

Ullrich, S.*, Nadelson, I.*, Krebs, S., Blum, H., Leonhardt, H., Solovei, I. (**2024**). Co-transcriptional splicing is delayed in highly expressed genes. BioRxiv. https://doi.org/10.1101/2024.06.04.597307

Li A., Yonchev I., **Ullrich S.,** Apostol C., Heath C., Griffith L., Daniels P., Whelan A., Solovei I, Sudbery I., Wilson S. A. hnRNPU restricts RNA polymerase II CTD phosphorylation and regulates transcription, mRNA processing and transcription loop size. ***Manuscript in preparation.***

Chudzik K., Trakarnphornsombat W., Sato Y., **Ullrich S**., Tooze R., Schermelleh L., Robson M., Solovei I., Kimura H. “Epi-jamming” creates artifactually peripheral signal in microscopy and genomics. ***Manuscript in preparation.***

Adams NM., Galitsyna A., Tiniakou I., Esteva E., Ra A., **Ullrich S**., Yeung S., Tan Y., Levy DE., Khanna KM., Mirny LA., Solovei I., Reizis B. Regulation of the type I interferon response by chromatin organization. ***Manuscript in preparation.***

Soni A., Klütsch D., **Ullrich S**., Rund N., Houtman J., Enzo S., Karasinsky A., Lesche M., Dahl A., Trimbuch T., Ehninger D., Poetsch A., Solovei I, Amin H. & Toda T. Nup153 regulates neuronal responsiveness through activity-regulated gene programs. ***Manuscript in preparation.***

***** These authors contributed equally

4.4 Declaration of contributions as a co-author

4.4.1 Contribution to "Spatial organization of transcribed eukaryotic genes"

I conducted multiple FISH stainings for *Tg*, *Ttn*, and *Cald1* across various experiments using oligo based SABER-FISH, as well as immunostainings for RNAPII S2p, S5p, and SON. I also performed transcription inhibition experiments using alpha-amanitin and actinomycin D, along with splicing inhibition experiments using pladienolide B. For all these experiments, I utilized confocal microscopy, prepared images and quantified the data. My work contributed to figures 2, 3, 5, and 7, as well as extended data figures 4, 5, and 6, and supplementary figures 3 and 5. Additionally, I contributed to the final manuscript both before submission and during revision.

4.4.2 Contribution to "Co-transcriptional splicing is delayed in highly expressed genes"

I analyzed nanopore RNA-seq data and conducted FISH stainings for *Tg*, *Cald1*, and individual introns of these genes. For all experiments, I utilized confocal microscopy and quantified data using the script provided by Iliya Nadelson. My work was included in figures 1, 2, 4, 5, 6, and 7, as well as supplementary figures 1, 2, and 3. Additionally, I contributed to the final manuscript before submission and during the revision process and assisted in creating the figures.

4.4.3 Contribution to "hnRNPU restricts RNA polymerase II CTD phosphorylation and regulates transcription, mRNA processing and transcription loop size"

I conducted siRNA knockdowns of hnRNPU in Pmi28 cells and performed immunostainings to assess the knockdown efficiencies of hnRNPU. Following this, I performed FISH for *Cald1* and *Ttn*. Using confocal microscopy, I captured images of the stained samples, wrote an image analysis pipeline and evaluated the resulting data. Additionally, I prepared images and graphs to be used in the manuscript.

4.4.4 Contribution to "'Epi-jamming' creates artifactually peripheral signal in microscopy and genomics"

I performed immunostainings utilizing multiple antibodies and nanobodies in Pmi28 and HeLa cells. I captured confocal images of the stained samples. Additionally, I developed an image analysis pipeline, and used it to evaluate the data.

4.4.5 Contribution to "Regulation of the type I interferon response by chromatin organization"

I performed immunostainings for IFN-α and the Golgi apparatus marker GORASP2 in pDCs. Additionally, I conducted FISH stainings for the IFN-I locus and a TAD containing the Slc24a2 gene in undifferentiated HoxB8-FL progenitor cells, unstimulated and stimulated pDCs, uninfected and virus-

infected fibroblasts, as well as Pmi28 cells. I scanned images and evaluated the data, mainly the distance between FISH signals. Furthermore, I prepared images and evaluation graphs to be used in figures.

4.4.6 Contribution to "Nup153 regulates neuronal responsiveness through activity-regulated gene programs"

I assisted Abhinav Soni in designing, ordering and preparing SABER-FISH probes for various candidate genes. Furthermore, I guided him on imaging the samples with a confocal microscope and analyzing the resulting data.

4.5 Declaration

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 08.01.2025 Simon Ullrich

Erklärung

Hiermit erkläre ich, *

- [X] dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

- [X] dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.

- [ ] ~~dass ich mich mit Erfolg der Doktorprüfung im Hauptfach ..... und in den Nebenfächern ..... bei der Fakultät für ..... der ..... unterzogen habe. (Hochschule/Universität)~~

- [ ] ~~dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.~~

München, den 08.01.2025 Simon Ullrich

*) Nichtzutreffendes streichen

4.6 Acknowledgements

First, I would like to express my sincere appreciation to Dr. Irina Solovei for her exceptional supervision and mentorship. Thank you for your support in my academic and personal endeavors. Your insightful feedback and encouragement have played a crucial role in shaping this thesis.

I am grateful to Prof. Heinrich Leonhardt for providing me with the opportunity to undertake this research project in his lab, as well as for his scientific support and guidance throughout this process. I also want to thank my TAC committee members Prof. Peter Becker, Dr. Boyan Bonev, and Dr. Sebastian Bultman for their helpful input and guidance during the project.

Furthermore, I wish to thank the members of the AG Leonhardt for creating a friendly atmosphere and providing lots of assistance during my - admittedly only occasional - visits to the BiosysM. I want to emphasize Jeanette's invaluable assistance in maintaining the smooth operation of the lab and handling organizational tasks. I would also like to express my gratitude to my former students Paul, Iliya, and Kyril for the time we spent together in the lab. I thoroughly enjoyed working with you.

Also a sincere thank you to the Cell Biology Department at the Biocenter, particularly Felix, Nupur, Ria, Simon, Mengqiao, Johannes, Nadja, Marcel, Nils, Shalini, and Soraya, for making my time inside and outside the lab so enjoyable, whether during work, lunch, our coffee breaks or in my leisure time.

Finally, I would like to extend my gratitude to my family - my parents, my sister, Maya, Maxie, and the newest addition to the family, Ben - for their steadfast support. None of this would have been possible without all of you!