

# THE SPATIAL ORGANIZATION OF TRANSCRIBED EUKARYOTIC GENES

Dissertation an der Fakultät für Biologie  
der Ludwig-Maximilians-Universität



vorgelegt von  
Susanne Leidescher

München  
April 2020

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

Erstgutachter: Prof. Dr. Heinrich Leonhardt  
Zweitgutachter: Prof. Dr. Dirk Eick  
Tag der Abgabe: 23.04.2020  
Tag der mündlichen Prüfung: 08.10.2020

## ERKLÄRUNG

Ich versichere hiermit an Eidesstatt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist. Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt. Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 23.04.2020

Susanne Leidescher

# TABLE OF CONTENTS

<b>Summary .....</b>	<b>1</b>
<b>Zusammenfassung .....</b>	<b>3</b>
<b>1 INTRODUCTION .....</b>	<b>5</b>
1.1 Hierarchical genome organization .....	5
1.1.1 Chromatin organization .....	5
1.1.2 Spatial segregation of eu- and heterochromatin .....	5
1.1.3 Chromosome territories (CTs) .....	8
1.1.4 Topologically associating domains (TADs) .....	9
1.1.5 Enhancer-promoter loops .....	9
1.2 The spatial organization of transcription .....	11
1.2.1 Early microscopic visualization of transcription .....	11
1.2.2 The transcription factory model and nuclear speckles .....	13
1.2.3 Transcriptional bursting .....	14
1.2.4 Phase separation .....	14
1.2.5 Light microscopy resolution as a limiting factor .....	15
1.3 Aims and scope of this study .....	16
<b>2 MATERIALS AND METHODS .....</b>	<b>17</b>
2.1 Cell culture .....	17
2.1.1 Culture of Pmi28 mouse myoblasts .....	17
2.1.2 Myotube differentiation .....	17
2.1.3 Poly-L-lysine coating of coverslips .....	17
2.1.4 Transcription inhibitor treatment .....	18
2.2 Tissue preparation .....	19
2.2.1 Isolation and fixation of mouse tissues .....	19
2.2.2 Thyroid tissue of other vertebrates .....	20
2.2.3 Embedding and cryosectioning of tissue .....	20
2.3 Transcription stimulation via dCas9-VPR .....	21
2.3.1 gRNA expression plasmid generation .....	21
2.3.2 Stable dCas9-VPR expression .....	21
2.3.3 Transfection of gRNA plasmids .....	22
2.3.4 Analysis .....	22
2.4 Fluorescence in situ hybridization (FISH) .....	23
2.4.1 Labeling of dUTPs .....	23
2.4.2 Preparation of dNTP mix for Nick Translation and DOP-PCR .....	24

2.4.3	Nick translation (NT) .....	24
2.4.4	FISH probe design and preparation .....	25
2.4.5	Preparation of hybridization mixture .....	28
2.4.6	2D FISH .....	29
2.4.7	3D FISH on tissue cryosections .....	29
2.4.8	3D FISH on adherently growing cells.....	30
2.4.9	ImmunoFISH.....	31
2.5	Gene expression analysis .....	33
2.5.1	RNA isolation .....	33
2.5.2	cDNA preparation and quantitative real time PCR (qPCR) .....	33
2.5.3	RNA sequencing (RNA-seq) .....	34
2.5.4	Nanopore sequencing (Nanopore-seq).....	35
2.6	Native Chromatin ImmunoPrecipitation followed by sequencing (nChIP-seq) .....	36
2.7	Chromatin Conformation Capture (Hi-C).....	39
2.8	Microscopy.....	40
2.8.1	Light microscopy .....	40
2.8.2	Electron microscopy.....	41
2.9	Bioinformatics .....	41
2.9.1	GTEx data analysis.....	41
2.9.2	Semi-automated gene flanks measurement .....	41
2.9.3	Computational chromatin modeling .....	42
<b>3</b>	<b>RESULTS.....</b>	<b>43</b>
3.1	Long mammalian genes are rarely highly expressed.....	43
3.1.1	The majority of mammalian genes is short .....	43
3.1.2	Long human genes are rarely highly expressed .....	44
3.1.3	Long genes exhibit lower polymerase levels than short genes .....	45
3.1.4	Selection of long highly expressed human genes from GTEx dataset .....	47
3.1.5	Validation of similar characteristics for orthologous mouse genes.....	47
3.2	Expressed long genes form transcription loops .....	49
3.2.1	FISH probe design.....	49
3.2.2	3D FISH on long highly expressed genes reveals transcription loops .....	49
3.2.3	Progression of transcription along TLs: Visualization of exons and elongating polymerases...52	
3.2.4	Progression of transcription along TLs: Visualization of introns and nuclear speckles .....	54
3.2.5	Absence of recursive splicing for the longest <i>Tg</i> intron.....	56
3.3	Induction and inhibition of transcription loops .....	57
3.3.1	Initiation of <i>Tm</i> transcription loops by transcription initiation .....	57
3.3.2	Elimination of <i>Tm</i> transcription loops by transcription inhibition.....	58

3.4	Correlation of transcription loop size, gene size and transcription level .....	60
3.4.1	Highly expressed genes are expressed biallelically.....	60
3.4.2	Transcription loop size does not correspond to gene size.....	60
3.5	Transcription loops shape their immediate chromatin neighborhood.....	63
3.5.1	<i>Tg</i> transcription loops are excluded from harboring chromosome territories.....	63
3.5.2	Hi-C data confirm exclusion of transcription loops from chromosomes .....	64
3.5.3	Transcription loops are open loops with separated flanks.....	66
3.6	Deciphering the cause of transcription loop formation.....	68
3.6.1	Are transcription loops stiff structures?.....	68
3.6.2	The longest intron of the <i>Tg</i> gene is strongly expanded.....	68
3.6.3	Electron microscopy reveals unique accumulations of granules in myotube nuclei.....	70
3.6.4	Polymer modeling confirms transcription loop stiffness .....	71
3.7	Taking a closer look at the extraordinary thyroglobulin gene.....	73
3.7.1	<i>Tg</i> forms transcription loops in other vertebrate species .....	73
3.7.2	<i>Tg</i> forms transcription loops in <i>in vitro</i> grown and in grafted thyroid follicles.....	75
3.7.3	<i>Tg</i> is highly upregulated and forms transcription loops upon birth.....	77
3.7.4	<i>Tg</i> is highly expressed and forms transcription loops irrespective of thyroid hormone levels ..	78
3.7.5	Mature <i>Tg</i> transcripts do not retain introns .....	81
<b>4</b>	<b>DISCUSSION AND FUTURE PERSPECTIVES .....</b>	<b>83</b>
4.1	Transcriptional bursting shapes transcription loop structure.....	84
4.2	Transcription loops are stiff structures due to dense decoration with ribonucleoprotein particles .....	87
4.3	Transcription loop formation as universal mode of transcription .....	89
<b>5</b>	<b>APPENDIX.....</b>	<b>91</b>
5.1	References.....	91
5.2	Supplementary tables .....	105
5.3	Abbreviations .....	116



---

## SUMMARY

Transcription is the first step on the complex way of converting genetic information into proteins. In recent years, various studies have contributed to a vast knowledge about eukaryotic transcription – on both the global nuclear and the molecular level. However, knowledge pertaining to an intermediate level of transcriptional organization, that is how individual expressed genes are spatially organized, is surprisingly limited. The only information about arrangements of transcribed chromosomal regions derives from studies of exceptional types of chromosomes, namely lampbrush chromosomes – meiotic bivalents in diplotene oocytes of most of vertebrates – or polytene chromosomes in salivary glands of Diptera larvae. Transcription units of these chromosomes – due to their prominent length and high expression level – form microscopically resolvable loops densely decorated by nascent RNA transcripts.

In striking contrast, expressed genes in interphase or post-mitotic somatic nuclei in mammals, as revealed by Fluorescence in situ Hybridization (FISH), do not form loops but exhibit a focus-like shape. This discrepancy can be attributed to (i) the limitation of light microscopy not allowing to resolve structures of short expressed genes and/or to (ii) a relatively low expression level of previously studied long genes.

Therefore, for this study, we first selected long highly expressed genes from the human GTEx consortium database, and confirmed the expression level for the orthologous mouse genes by RNA-seq. We selected five genes: thyroglobulin (*Tg*) expressed in thyroid secretory cells, titin (*Ttn*) and nebulin (*Neb*), both expressed in striated muscles and cultured myotubes, as well as myosin heavy chain (*Myh11*) and caldesmon (*Cald1*), both expressed in smooth muscles. Using RNA and DNA FISH, we visualized these genes and found that, upon expression, they form huge loops. These loops are characterized by ongoing transcription and structurally highly similar to the loops formed by transcription units of lampbrush and polytene chromosomes. Thus, we termed these loops “transcription loops” (TLs).

We further confirmed transcription-dependent dynamics of TL formation by experimentally manipulating transcription levels of the *Ttn* gene, using inhibitory drugs and a CRISPR/Cas9-based approach. A combination of FISH and Hi-C analysis allowed us to show that TLs dynamically shape harboring chromosomal loci: TLs protrude into the nuclear space, separate gene flanking regions and even modify the structure of their corresponding chromosome territories. These features of TLs suggested that they are intrinsically stiff structures. Based on our observations, we hypothesized that TL stiffness is caused by the dense decoration of a highly expressed gene with ribonucleoprotein particles, which are formed concomitantly with transcription progression. Three lines of evidence, obtained by (i) differential labeling of nascent RNA along a gene body, (ii) electron microscopy of myotube nuclei and (iii) polymer modeling of *Tg* TLs, supported our hypothesis.

Taken together, our results indicate that TL formation is a universal principle of eukaryotic transcription. Our work sets the ground for reconsiderations of previous assumptions on the spatial organization of transcribed genes in eukaryotes.





---

## ZUSAMMENFASSUNG

Transkription ist der erste Schritt auf dem komplexen Weg von der genetischen Information zum fertigen Protein. Viele Studien – sowohl den Zellkern im Ganzen betreffend als auch auf molekularer Ebene – haben zu einem besseren Verständnis der eukaryotischen Transkription beigetragen. Erstaunlicherweise ist jedoch das Wissen über die Zwischenebene der transkriptionellen Organisation, nämlich die räumliche Struktur einzelner exprimierter Gene, sehr begrenzt. Die einzigen Informationen über die räumliche Organisation exprimierter Loci finden sich in Studien, die an zwei Arten von außergewöhnlichen Chromosomen durchgeführt wurden: zum einen an sogenannten “lampbrush” Chromosomen, meiotischen Bivalenten in diplotenen Oozyten der meisten Wirbeltiere, und zum anderen an polytenen Chromosomen in den Speicheldrüsen von Larven der Ordnung Diptera. Aufgrund ihrer enormen Länge und ihres hohen Expressionslevels, bilden Transkriptionseinheiten in diesen Chromosomen mikroskopisch auflösbare Schleifen, die dicht mit RNA bepackt sind.

Im deutlichen Gegensatz dazu bilden exprimierte Gene in post-mitotischen Säugerzellkernen oder in Säugerzellkernen, die sich in der Interphase befinden, keine Schleifen, sondern kleine Foci. Dieser Unterschied lässt sich auf zwei Arten erklären: (i) durch die lichtmikroskopische Auflösungsgrenze, welche die Auflösung der Struktur kurzer exprimierter Gene verhindert und/oder (ii) durch die verhältnismäßig niedrige Expression von langen untersuchten Genen.

Folglich haben wir für unsere Studie zunächst lange und hoch exprimierte Gene aus der humanen GTEx Datenbank ausgewählt und die Expressionslevel der orthologen Gene in den entsprechenden Mausgeweben mittels RNA-seq bestätigt. Die Auswahl umfasste fünf Gene: Thyroglobulin (*Tg*), das in sekretorischen Zellen der Schilddrüse exprimiert wird, Titin (*Ttn*) und Nebulin (*Neb*), die beide spezifisch in Skelettmuskelzellen und Myotubenkulturen exprimiert werden, sowie eine Myosinkomponente (*Myh11*) und Caldesmon (*Cald1*), die beide in Muskelzellen der glatten Muskulatur exprimiert werden. Mittels DNA- und RNA-Fluoreszenz in situ Hybridisierung (FISH), machten wir diese Gene sichtbar und zeigten, dass sie im aktiven Zustand große Schleifen bilden. Diese Schleifen sind durch laufende Transkription gekennzeichnet und strukturell sehr ähnlich zu den Schleifen, welche man in lampbrush Chromosomen und polytenen Chromosomen findet. Folglich bezeichneten wir diese Schleifen als Transkriptionsschleifen („Transcription loops”).

Des Weiteren bestätigten wir die transkriptionsabhängige Dynamik der Schleifenbildung, indem wir die Expressionsrate des *Ttn* Gens mit Transkriptionseinhibitoren und mit einem CRISPR/Cas9 basierten Ansatz experimentell beeinflussten. Mit einer Kombination aus FISH und Hi-C konnten wir zeigen, dass Transkriptionsschleifen ihre direkte chromosomale Umgebung dynamisch formen: Transkriptionsschleifen ragen ins Nucleoplasma, führen zur Trennung der unmittelbar an das betreffende Gen angrenzenden Regionen und verändern sogar die Struktur des jeweiligen Chromosomenterritoriums. Diese Eigenschaften der Transkriptionsschleifen legten nahe, dass es sich bei ihnen um Strukturen mit inhärenter Steifheit handelt. Auf Grundlage dieser Beobachtungen vermuteten wir, dass die Steifheit durch die dichte Bepackung eines exprimierten Gens mit Ribonucleoproteinkomplexen, die im Zuge der Transkription entstehen, verursacht wird. Drei Beobachtungen, basierend auf (i) der Visualisierung von RNA an bestimmten Teilen eines Gens,

(ii) elektronenmikroskopischen Untersuchungen von Myotubezellkernen und (iii) Polymermodellen der *Tg* Transkriptionsschleife, unterstützen unsere Hypothese.

Zusammenfassend deuten unsere Ergebnisse darauf hin, dass die Bildung von Transkriptionsschleifen ein universelles Prinzip der eukarotischen Transkription darstellt. Unsere Arbeit legt die Grundlage für eine Neubetrachtung bestehender Annahmen bezüglich der strukturellen Organisation transkribierter eukaryotischer Gene.

---

# 1 INTRODUCTION

## 1.1 HIERARCHICAL GENOME ORGANIZATION

### 1.1.1 CHROMATIN ORGANIZATION

In eukaryotic cells, the genome is separated from the cytoplasm and confined to the nucleus. To achieve this, the genomic DNA has to be fairly tightly packed. Therefore, it is complexed with various proteins and RNA into a structure referred to as chromatin (Kornberg, 1974). The basic structural unit of chromatin is a nucleosome, consisting of a histone octamer and 147 base pairs (bp) of DNA wrapped around it (Luger et al., 1997). The canonical histone octamer comprises two copies of each of the histones H2A, H2B, H3 and H4 (Luger et al., 1997). Nucleosomes measure 11 nm in diameter and are arranged into a linear array spaced by a linker region of 10 to 50 bp, resembling a “beads on a string” structure (Baldi et al., 2020; Olins and Olins, 1974; Olins and Olins, 2003). The linker region provides a binding platform for histone H1 or other non-histone proteins.

During development, throughout the cell cycle or upon the impact of a stimulus, the chromatin environment undergoes dramatic changes. These changes are not based on alterations in the underlying DNA sequence but on marks set on chromatin and represent the epigenetic (*epi*: Greek for *above, on top of*) level of chromatin regulation (Waddington, 1942). The best studied epigenetic marks include DNA methylation, post-translational modifications (PTMs) of histone tails, incorporation of histone variants and non-coding RNAs (ncRNA). All of these marks regulate transcriptional output and shape cellular identity by altering chromatin accessibility or serving as interaction platform for additional regulatory factors. In recent years, evidence emerged that not only the described factors but also the three dimensional arrangement of chromatin within the nucleus itself strongly influences transcriptional output and therefore controls genome function (see below).

### 1.1.2 SPATIAL SEGREGATION OF EU- AND HETEROCHROMATIN

The tight orchestration of spatio-temporal gene expression patterns is crucial for the correct development of an organism (Reik, 2007; Zheng and Xie, 2019). One of the main requirements to achieve this orchestration is the spatial separation of active and inactive chromatin (Kosak et al., 2007; Misteli, 2007). Genes that are actively transcribed reside in the so-called euchromatin (EC) in the nuclear interior whereas silent genes are sequestered to heterochromatin (HC) at the nuclear periphery (Solovei et al., 2016; van Steensel and Belmont, 2017) (Figure 1.1 A). Eu- and heterochromatin not only differ in the level of gene activity but also in their sequence composition and associated chromatin binding factors as well as epigenetic marks (Bannister and Kouzarides, 2011; Bernardi, 1989; Jenuwein and Allis, 2001; Kornberg and Rykowski, 1988). Euchromatin is decondensed (“open”) and accessible to binding of regulatory factors. It is generally characterized by a high

GC content and the presence of active histone modifications such as H3K4me3 and H3K9ac at promoters and H3K36me3 at gene bodies (Bannister et al., 2005; Barski et al., 2007; Liang et al., 2004). Heterochromatin, in contrast, is highly condensed (“closed”), rather inaccessible for binding factors, and is defined by a low GC content as well as repressive histone modifications. HC comprises two different forms – facultative HC (fHC) and constitutive HC (cHC) – that differ in their transcriptional activity as well as epigenetic marks. fHC contains tissue-specific and differentially expressed genes that are rapidly up- or downregulated at a certain developmental stage. fHC is further marked by the repressive histone modifications H3K27me3, H3K9me2, and H4K20me1 (Peters et al., 2003). The inactive X chromosome ensuring dosage compensation in mammalian female cells serves as an example of facultative heterochromatin: it is characterized by H3K27me3 marks and association with polycomb repressive complexes (PRCs) (Dossin et al., 2020; Trojer and Reinberg, 2007).

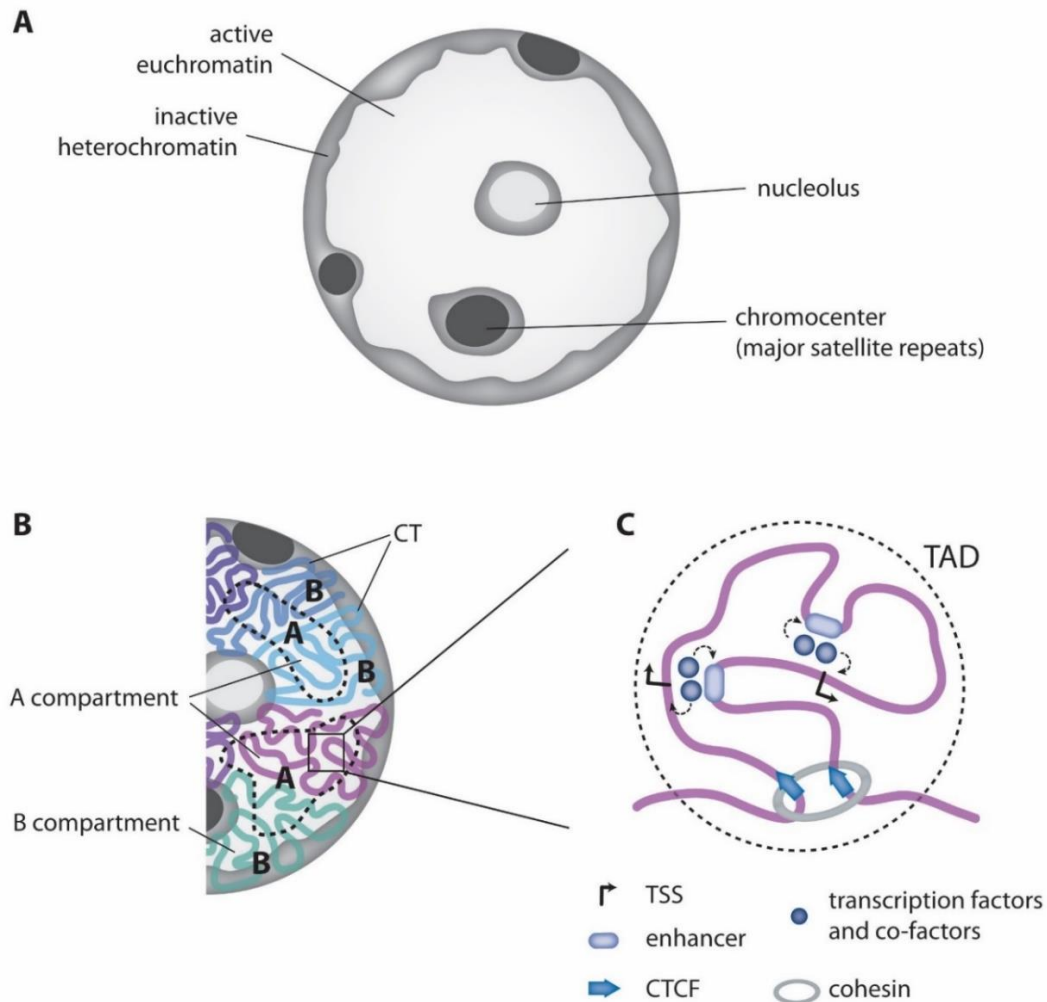
cHC predominantly consists of highly repeated sequences (satellite DNA) accumulated mostly in sub-centromeric and sub-telomeric regions and remains condensed throughout development (Grewal and Jia, 2007). It is characterized by the repressive histone marks H3K27me1, H4K20me3, and H3K9me3 (Peters et al., 2003; Schotta et al., 2004). H3K9me3 serves as interaction platform for heterochromatin protein 1 (HP1), a family of three proteins crucial for heterochromatin formation via a positive feedback loop propagating the H3K9me3 mark along repressed chromatin by interaction with the H3K9 methyltransferase Suv39 (Bannister et al., 2001; Maison and Almouzni, 2004). In mouse cell nuclei, the cHC at pericentromeric regions of acrocentric chromosomes clusters to form dense spherical structures called chromocenters (CCs) intensely stained by 4',6-diamidino-2-phenylindole (DAPI) (Guenatri et al., 2004) (Figure 1.1 A).

Genome wide chromosome conformation capture (Hi-C) analysis of mammalian genomes showed that chromatin can be divided into specific multi mega base pair (Mb) domains, termed A and B compartments. A and B compartments correspond to EC or HC, respectively (Lieberman-Aiden et al., 2009; Zhang et al., 2012) (Figure 1.1 B). Gene-rich, transcriptionally active A compartments cluster preferentially with other A compartments and gene-poor, transcriptionally silent B compartments cluster preferentially with B compartments, but do not mix (Gibcus and Dekker, 2013).

An exception from the conventional nuclear architecture with EC residing in the nuclear interior and HC on the nuclear periphery was found in the nuclei of rod photoreceptor cells of nocturnal mammals. In these nuclei, the conventional genome architecture is inverted and a dense heterochromatic nuclear core is surrounded by peripheral euchromatin. This inversion facilitates the vision of nocturnal mammals by channeling the light through the dense nuclear cores and thereby reducing light scattering in the retina (Solovei et al., 2009).

Polymer modeling approaches in combination with Hi-C and Fluorescence in situ Hybridization (FISH) revealed the driving force for spatial separation of eu- and heterochromatin in mammalian nuclei. Interestingly, euchromatic interactions are dispensable for the formation of chromatin compartments, whereas heterochromatic interactions are essential for this formation in both, conventional and inverted

nuclei. For building the conventional chromatin distribution, tethering of heterochromatin to the lamina needs to be established, suggesting that without such tether all nuclei would be inverted (Falk et al., 2019).



**FIGURE 1.1: Hierarchical genome organization. (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 and chromocenters (dark gray). **(B)** Each chromosome (purple, dark blue, light blue, magenta, and turquoise) occupies a distinct chromosome territory (CT). Chromosome folding within their territories leads to gene-rich, active regions residing in the euchromatic A compartment and gene-poor, inactive regions residing in the heterochromatic B compartment. Depending on the transcriptional status of a chromosomal region, the region shifts to the corresponding compartment. **(C)** Topologically associating domains (TADs) define regions of high long-range interaction frequencies (e.g. promoter enhancer loops) and are confined by boundaries characterized by architectural proteins such as cohesin and CTCF. TSS: transcription start site; CTCF: CCCTC-binding factor.

### 1.1.3 CHROMOSOME TERRITORIES (CTs)

Towards the end of the 20<sup>th</sup> century, researchers extensively investigated the spatial intra-nuclear distribution of chromosomes. Chromosome painting (Lichter et al., 1988; Pinkel et al., 1988) revealed that each eukaryotic chromosome occupies a distinct area in the nucleus, a so-called chromosome territory (CT) (Cremer and Cremer, 2010) (Figure 1.1 B). This finding was later confirmed by Hi-C studies (Lieberman-Aiden et al., 2009; Zhang et al., 2012). Importantly, CTs persist during interphase, but their spatial arrangement may vary from cell to cell in a population as well as during development or in different cell types (Mayer et al., 2005). Nevertheless, the distribution of chromosomes within the nucleus is not random. This holds true for all human chromosomes, with gene-poor late replicating chromosomes, such as HSA18 (*Homo sapiens* chromosome 18), localizing at the nuclear periphery and gene-rich, early replicating chromosomes, like HSA19 (*Homo sapiens* chromosome 19), residing in the nuclear interior (Boyle et al., 2001; Cremer et al., 2001; Croft et al., 1999; Habermann et al., 2001). The same principle of chromosome arrangement is observed in various species, including birds (Habermann et al., 2001; Tanabe et al., 2002), rodents (Mayer et al., 2005; Neusser et al., 2007), and cattle (Koehler et al., 2009). Not only CTs themselves, but also their sub-regions, follow the rule of radial distribution: gene-rich EC regions localize towards the nuclear interior, whereas gene-poor HC regions localize towards the nuclear periphery (Goetze et al., 2007; Kupper et al., 2007). Therefore, since EC and HC regions alternate along chromosomes, chromosomes fold between nuclear periphery and nuclear interior. Such folding is facilitated by so-called lamina-associated domains (LADs). These are regions that are bound to the nuclear lamina and alternate with so called inter-LADs, regions rarely contacting the lamina (Guelen et al., 2008). LADs are a major constituent of heterochromatic B compartments as identified by DNA adenine methyltransferase identification (DamID) (Guelen et al., 2008; van Steensel and Belmont, 2017).

Accumulating evidence suggests that chromosomal territoriality is dispensable for genome function within the nucleus and rather represents a consequence of mitotic division. Even though in mitosis partitioning into chromosomes is crucial to ensure proper genome division, chromosomes are microscopically indistinguishable thereafter (Solovei et al., 2016). Nevertheless, researchers still refer to CTs as a basic structural element of the nucleus.

Chromosomes are largely immobile during interphase, however, their sub-regions can be highly dynamic. Short-range movements (0.5 – 1  $\mu$ m) occur due to the inert constrained diffusion of chromatin, but also larger movements on the scale of a few microns have been observed, mostly linked to gene activation or silencing and taking place rather over cell division than during interphase (Chuang et al., 2006; Lanctot et al., 2007; Soutoglou and Misteli, 2007). Genes move within their respective CT depending on the transcriptional state, with expressed genes moving towards the borders of a territory and even looping out of it (Shah et al., 2018; Volpi et al., 2000).

#### 1.1.4 TOPOLOGICALLY ASSOCIATING DOMAINS (TADs)

Topologically associating domains (TADs) are formed by loci within the same compartment that are engaged in long-range contacts (Dixon et al., 2012; Hou et al., 2012; Sexton et al., 2012) (Figure 1.1 C) and include chromatin with similar expression level and replication timing. TADs were identified in humans (Dixon et al., 2012), mice (Dixon et al., 2012; Nora et al., 2012), and *Drosophila* (Sexton et al., 2012) by mapping chromatin interaction frequencies via Hi-C. In mice, the size of these domains ranges from tens of kilo base pairs (kb) to several mega base pairs (Mb), with the median being 880 kb (Dixon et al., 2012).

Genes within the same TAD are mostly co-regulated during differentiation and development (Nora et al., 2012) and the borders of TADs correspond to those of replication domains (Pope et al., 2014). Perturbing TAD insulation leads to severe misregulation of gene expression due to contacts of gene promoters with “false” regulatory elements resulting in developmental defects and cancer (Flavahan et al., 2016; Hnisz et al., 2016; Lupianez et al., 2015).

TAD formation has been recently attributed to chromatin loop extrusion by the cohesin complex (Nuebler et al., 2018). Besides the cohesion complex, TAD boundaries are predominantly associated with the CCCTC-binding factor CTCF (Phillips-Cremins et al., 2013; Vietri Rudan et al., 2015) (Figure 1.1 C). Notably, the orientation in which a CTCF site is present in the genome influences the formation of TAD boundaries – with TADs mostly forming between convergent CTCF sites (de Wit et al., 2015). A change in orientation or complete removal of CTCF sites leads to a TAD’s disappearance, shifting or fusion (Guo et al., 2015; Lupianez et al., 2015).

#### 1.1.5 ENHANCER-PROMOTER LOOPS

The regulation of mammalian transcription on a small scale requires the precise interplay of various factors, including transcription factors (TFs), the histone code and Pol II C-terminal domain (CTD) code, DNA modifications, and non-coding RNAs. To transcribe a gene in the correct spatio-temporal manner, crosstalk between the gene’s promoter and its non-coding regulatory elements is crucial. Depending on whether these elements interact intramolecularly (enhancers, silencers and insulators) or intermolecularly (transcription factors, lncRNA), they are referred to as *cis*- or *trans*-acting, respectively. One of the best-studied *cis*-regulatory elements are enhancers (Figure 1.1 C). Enhancers are short nucleotide sequences within the genome, which may be located a few kb to 1 Mb apart from their target promoter(s) and may reside in intra- or intergenic regions, including gene introns and exons (Birnbaum et al., 2012; Mifsud et al., 2015; Sanyal et al., 2012; Schoenfelder et al., 2015). These elements are characterized by histone modifications including H3K4me1, H3K27me3, H3K27ac, H3K36me3 and combinations thereof and thus influence the developmental state (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Schoenfelder and Fraser, 2019; Zentner et al., 2011). The diverse locations and different features of the regions make it very difficult to map putative enhancers (Schoenfelder and Fraser, 2019).

To stimulate the expression of its target gene, an enhancer region is bound by TFs and co-factors (Calo and Wysocka, 2013) and thus increases the activity of a promoter that resides in close proximity. As accumulating evidence suggests, the contact between a promoter and an enhancer is achieved by the formation of promoter-enhancer loops that are regulated by the Mediator initiation core complex (Nolis et al., 2009; Tolhuis et al., 2002; Williamson et al., 2016). Therefore, especially chromosome conformation capture techniques (C-techniques) like 4C or capture Hi-C that enable the identification of interactions between one locus and many or all other loci, respectively, together with dedicated computational pipelines, have been proven useful in locating putative enhancers through identifying promoter loops (Ron et al., 2017).

Enhancers vastly outnumber protein-coding genes as hundreds of thousands of putative enhancers have been mapped in the mouse genome (Andersson et al., 2014; Schoenfelder and Fraser, 2019; Shen et al., 2012). Therefore, the promoter of one gene is mostly influenced by more than one enhancer, augmenting the level of transcriptional regulation by additive enhancer action. However, the interaction of a promoter with an enhancer mainly occurs within TAD boundaries and transcription is correlated with local chromatin insulation (Bonev et al., 2017).

Clusters of enhancers in conjunction with a bulk of bound transcription factors are called super-enhancers (SEs) (Pott and Lieb, 2015). These SEs are found throughout the genome and are linked to the rapid upregulation of developmental genes to ensure cell identity and the transcriptional control of genes with prominent roles in cell-type specific processes (e.g. pluripotency genes *Oct4*, *Sox2*, and *Nanog* in embryonic stem cells (ESCs)) (Whyte et al., 2013). Furthermore, SEs are involved in the regulation of oncogenes in cancer cells (Hnisz et al., 2013). Although exact mechanisms leading to the attraction between promoters and enhancers, as well as to the formation of super-enhancers, are not known, it was suggested that liquid-liquid phase separation driven by proteins binding these genomic regions may play a crucial role in these processes (see also section 1.2.4).



## 1.2 THE SPATIAL ORGANIZATION OF TRANSCRIPTION

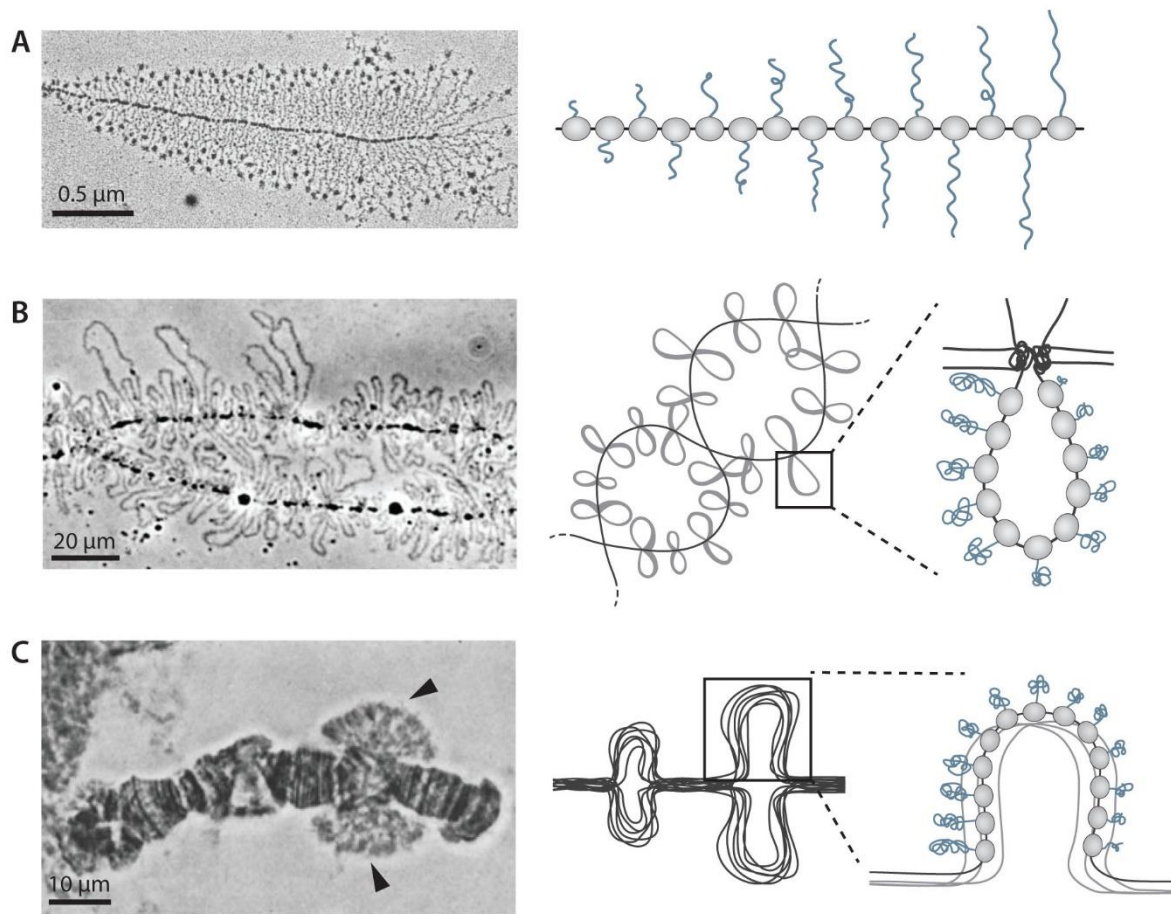
### 1.2.1 EARLY MICROSCOPIC VISUALIZATION OF TRANSCRIPTION

In the 1950s and 60s, a series of studies fundamentally contributed to the view on the spatial organization of transcription. The most important ones were conducted on Miller spreads of chromatin from various types of nuclei, lampbrush chromosomes (LBCs) isolated from amphibian diplotene oocytes, and polytene chromosomes of Diptera. Due to their size of several tens of microns and their distinctive structures, these types of chromosomes are perfectly suited for microscopic studies.

In 1969, Miller and Beatty invented an important method to visualize chromatin. This method, now commonly referred to as “Miller spreading” (Miller and Beatty, 1969a, b), is based on treating chromatin with very low ionic strength and alkaline pH to gently unfold it and preparing spreads on electron microscope (EM) grids. These spreads enabled researchers to visualize interphase chromatin and, in particular, observe transcription units with attached nascent RNA transcripts (Figure 1.2 A).

LBCs are diplotene bivalents formed in the course of meiosis in the oocytes of most animals except mammals (Macgregor, 1993). On these chromosomes, transcription units are highly active to ensure sufficient protein synthesis during oogenesis and early embryogenesis. The high expression of these long transcription units, in most cases exceeding the length of one gene, leads to the formation of chromatin loops emanating laterally from the chromatid axis, thus causing the bivalents to appear as fluffy brushes (Callan, 1986) (Figure 1.2 B). Miller spreads of LBCs suggested that on highly transcribed genes, the chromatin thread is massively stretched and depleted of nucleosomes (Franke et al., 1976).

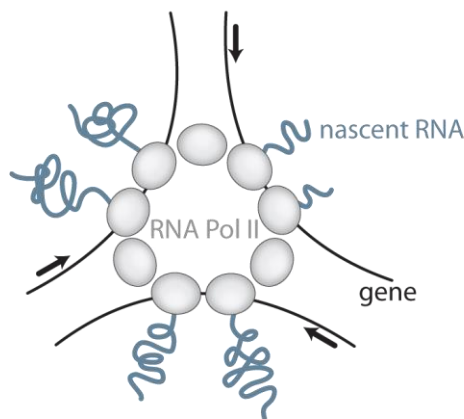
Polytene (“multi-stranded”) chromosomes arise from successive chromatid duplications in the absence of cell division (“endoreplication”) and consist of thousands of homologous DNA strands tightly attached to each other (Bjork and Wieslander, 2015). They are found in the interphase nuclei of several fly tissues but especially prominent in larval salivary glands (Macgregor, 1993). As in LBCs, highly transcribed genes loop out from the chromosome axis, but since thousands of strands lie next to one another, they do not form single loops but several collateral loops called “puffs”, the biggest of them known as “Balbiani rings” (Daneholt, 1975) (Figure 1.2 C). In both LBCs and polytene chromosomes, the key features for the microscopic visualization of transcription are (i) the immense size of the chromosomes and their transcription units and (ii) the high expression level of the visualized genes.



**FIGURE 1.2: Models for the early microscopic visualization of transcription.** **(A)** Left: Miller spreads prepared from rRNA genes of a *Pleurodeles* oocyte show a flattened gene axis with nascent RNAs emerging from it. RNAs grow in size as transcription progresses (from left to right) and lead to the appearance of a “christmas-tree” like shape characteristic for this type of chromatin preparation. Adapted from (Scheer, 1987). Right: Schematic depiction of the gene axis (black line) on a Miller spread covered by elongating polymerases (gray) with attached nascent transcripts (blue). **(B)** Left: Phase contrast micrograph of a part of a lampbrush chromosome. The bivalent chromosome axes are covered by lateral loops formed by highly expressed transcription units. Adapted from (Macgregor, 2012). Right: Schematic depiction of transcription on lampbrush chromosomes. Transcription units are covered by polymerases (gray) that produce nascent RNA (blue). Increasing transcript length leads to the formation of an RNA gradient on the lateral loops. **(C)** Left: Polytene chromosome from a *Chironomus* salivary gland. Regions with increased transcription loop out from the chromosome axis and form so-called “puffs”. Especially single copy genes encoding the largest proteins in salivary gland cells are highly expressed and form so-called “Balbiani rings” (black arrowhead). Adapted from (Diez, 1973). Right: Schematic depiction of transcription on Balbiani rings. Thousands of DNA strands lie next to one another as a result of endoreplication. Since the DNA strands are homologous, increased transcription (and therefore dense loading with polymerases (gray) and nascent transcripts (blue)) of certain genes accumulates within the same region in all strands and leads to looping out of these regions from the chromosome axis.

### 1.2.2 THE TRANSCRIPTION FACTORY MODEL AND NUCLEAR SPECKLES

Even though knowledge about transcriptional fine regulation as well as the three-dimensional organization of the mammalian genome on a broad scale is largely comprehensive, the intermediate level of transcriptional organization is poorly understood. Based on the microscopic visualization of actively transcribed genes, elongating polymerases and nascent RNA transcripts, the existence of transcription factories has been proposed (Ghamari et al., 2013; Iborra et al., 1996; Jackson et al., 1993; Papantonis and Cook, 2013; Wansink et al., 1993). Transcription factories are distinct loci distributed throughout the nucleoplasm, at which hyperphosphorylated RNA polymerases are immobilized in groups (presumably by anchoring to a nuclear matrix). Upon transcription stimulation, genes approach these factories and are then reeled through the polymerases. Therefore, the extruding nascent RNAs are confined to a small volume, a single focus (Figure 1.3).



**FIGURE 1.3: Schematic depiction of a transcription factory.** Genes to be transcribed are reeled through immobilized polymerases. Thereby, nascent mRNA is produced at small discrete spots. Co-regulated genes approach a transcription factory at the same time and are transcribed simultaneously.

Supposedly, 8-30 RNA polymerases accumulate in a single transcription factory (Jackson et al., 1998; Martin and Pombo, 2003) with a diameter of about 75 nm on average, as determined by electron microscopy (Iborra et al., 1996). However, sizes may range from 40 to 200 nm depending on the associated factors (Eskiw and Fraser, 2011; Eskiw et al., 2008), linking the size of a factory to transcriptional activity, with many highly transcribed genes recruiting more polymerases (Rieder et al., 2012). The number of transcription foci per cell depends on the cell type and its differentiation state and varies between a few hundred and >1500 (Jackson et al., 1998; Osborne et al., 2004). Accumulation of genes in transcription factories might facilitate the quick and coordinated expression of co-regulated genes during differentiation or in response to cellular stimuli (Osborne et al., 2004).

Foci with RNA Pol II and nascent RNA accumulation have been shown to localize with or in close proximity to domains formed by splicing factor SC35 (Jackson et al., 1993; Wansink et al., 1993), a major constituent of nuclear speckles, indicating that co-transcriptional splicing occurs directly at these sites.

This co-localization is especially prominent for highly transcribed genes found in close proximity to nuclear speckles (Chen et al., 2018b; Hu et al., 2009; Khanna et al., 2014; Shopland et al., 2003). Moreover, it was shown that co-regulated and co-transcribed genes associate rather with the same nuclear speckle than with the same Pol II accumulation (Brown et al., 2008), challenging the transcription factory hypothesis.

### 1.2.3 TRANSCRIPTIONAL BURSTING

Labeling nascent RNAs enables the visualization of transcription kinetics. The visualization of RNAs *in vivo* mostly relies on RNA binding proteins fused to a fluorescent protein. For instance, in the MS2-MCP system, an array of stem loops derived from the MS2 bacteriophage is incorporated into the RNA of interest and then detected by the phage derived MS2 coat protein (MCP) fused to a fluorescent protein (Bertrand et al., 1998; Grunwald and Singer, 2010; Johansson et al., 1997; Larson et al., 2011; Peabody, 1993; Wu et al., 2016). Studies visualizing transcripts in prokaryotes (Cai et al., 2006; Golding et al., 2005), yeast (Tan and van Oudenaarden, 2010; Zenklusen et al., 2008) and other eukaryotes (Raj et al., 2006; Raj et al., 2010; Suter et al., 2011) have shown that transcription is a discontinuous process. Transcription occurs in so-called bursts, characterized by alternating active (ON) and inactive (OFF) states of a promoter (Chubb et al., 2006; Raj et al., 2006). Fluorescent experimental read-outs as described for the MS2/MCP system detect ON and OFF promoter states and enable to measure their frequency, amplitude and duration.

Using this system to visualize HIV-1 RNA, polymerase complexes were shown to move along a transcribed gene in groups termed convoys. A Pol-II convoy consists of closely spaced polymerases, and – in case of HIV-1 – includes on average approx. 20 enzyme complexes (Tantale et al., 2016). Moreover, factors like transcription factor availability, DNA looping and the composition of the local chromatin environment regulate different aspects of transcriptional bursting (Nicolas et al., 2017). Perturbations of the transcriptional co-activator Mediator severely influence the formation of RNA Pol II convoys, highlighting its prominent role in transcriptional regulation. For instance, knock-down of the Mediator subunit MED11 led to a reduction of polymerases per convoy, an increase of the spacing between the polymerases, and a shorter overall promoter ON time (Tantale et al., 2016).

### 1.2.4 PHASE SEPARATION

Live cell super-resolution studies on the spatio-temporal organization of transcription foci revealed that they are transiently formed (Buckley and Lis, 2014; Cisse et al., 2013) and contain proteins with intrinsically disordered domains, such as the activation domains of TFs and the C-terminal domain (CTD) of RNA Pol II (Boehning et al., 2018; Chong et al., 2018). These intrinsically disordered domains favor the formation of a separate liquid phase within the nucleoplasm. Since this phase is enriched in intrinsically disordered molecules, it separates from other liquid phases that are depleted of these molecules. This process is known as liquid-liquid phase separation (LLPS) (Hyman et al., 2014). Due to their disordered activation domains, TFs can promote phase separation – which is in turn associated with gene activation (Boija et al., 2018).

Additionally, TFs can recruit the intrinsically disordered CTD of Pol II and the CTD can also phase separate on its own (Boehning et al., 2018; Kwon et al., 2013). Therefore, formation of transient transcription clusters is likely dependent on LLPS of proteins with disordered regions (Hnisz et al., 2017).

The various transcription-coupled phase separating condensates differ in their composition (Cramer, 2019). Whereas the transcriptional co-activator Mediator and RNA Pol II are found in condensates on sites of active transcription (Cho et al., 2018), condensates found at super enhancers are enriched in Mediator and BRD4 (Sabari et al., 2018). Two other forms of transient condensates have been determined by microscopy: one at promoters and one at gene bodies with RNA Pol II shuttling between them (Cramer, 2019). At promoters, Pol II forms condensates with TFs, co-activators, initiation factors, and unphosphorylated Pol II. When the CTD of RNA Pol II is phosphorylated via cyclin dependent kinase 7 (CDK7) to promote elongation, it no longer forms a phase with TFs but now with a disordered region in the positive transcription elongation factor P-TEFb (Lu et al., 2018). At this point, the condensate loses its promoter specificity and forms an elongation-dependent gene body condensate containing phosphorylated Pol II, nascent RNA, elongation factors, and RNA processing factors (Cramer, 2019). These condensates in turn might correspond to foci where Pol II and splicing factors co-localize (Herzel et al., 2017; Misteli and Spector, 1999; Mortillaro et al., 1996).

### 1.2.5 LIGHT MICROSCOPY RESOLUTION AS A LIMITING FACTOR

Even though the development of C-techniques vastly improved our understanding of transcription-dependent genome organization, microscopic visualization of the structure of transcribed genes in cell nuclei is still elusive. The main constraint in this regard is the resolution limit of light microscopy (Schermerle et al., 2010). Indeed, a gene with the size of 10 kb would measure approx. 0.5  $\mu\text{m}$  if the chromatin fiber were stretched. However, genes as parts of chromosomal regions are rather coiled in the nucleus and therefore should be even smaller. These considerations tentatively explain why visualizations of genes by FISH result in dot-like signals with irresolvable 3D structure.

### 1.3 AIMS AND SCOPE OF THIS STUDY

In recent years, we have gained a vast knowledge about eukaryotic transcription – on both global nuclear and molecular levels. Paradoxically, knowledge pertaining to an intermediate level of transcriptional organization, that is how individual expressed genes are spatially organized, is surprisingly limited. The only information about the arrangement of transcribed genes is achieved from studies of meiotic lampbrush chromosomes in diplotene oocytes in vertebrates and polytene chromosomes in salivary glands of Diptera. In both cases, due to their length and expression level, genes and transcription units form microscopically resolvable loops densely decorated by nascent RNA transcripts. In contrast, expressed genes in mammalian interphase nuclei are resolved as foci restricted to a small nuclear area. The discrepancy between these observations could be attributed to (i) the limitation of light microscopy, not allowing to resolve genes that are coiled within the nucleus and/or to (ii) a relatively low expression level of the studied genes, when compared to the studied genes in lampbrushes and polytenes.

In this thesis, we set out to gain deeper insights into the spatial organization of transcription. Based on observations on LBCs and polytene chromosomes, we hypothesized that all expressed genes form loops and test this hypothesis by investigating the appearance of long and highly expressed genes utilizing a combination of imaging and molecular techniques.

First, we selected candidate genes that are long and highly expressed, and microscopically analyzed their appearance in both silent and expressed transcriptional states. Second, we experimentally manipulated the expression level of the genes and tested their morphology upon irreversible or reversible transcription inhibition, as well as after ectopically induced transcription. Third, we set out to determine what causes genes to exhibit a certain morphology and finally propose a model for the spatial organization of transcribed genes.

---

## 2 MATERIALS AND METHODS

### 2.1 CELL CULTURE

#### 2.1.1 CULTURE OF PMI28 MOUSE MYOBLASTS

Pmi28 cells (kindly donated by Prof. Dr. Cristina Cardoso, Technical University Darmstadt, Darmstadt, Germany) were routinely cultivated in Nutrient Mixture F-10 Ham (Sigma-Aldrich) supplemented with 20 % FCS (Biochrom) and 1 % Penicillin/Streptomycin (Sigma-Aldrich) at 37 °C and 5 % CO<sub>2</sub>. Every 2 to 3 days, or before cells reach confluency, they were passaged as follows: cells were trypsinized (1x Trypsin-EDTA solution, Sigma-Aldrich) for 3 min at 37 °C and resuspended in fresh medium. A fraction of cells (usually 1/5 or 1/10, depending on cell density) was transferred into a new tissue culture flask for further differentiation. The remaining cells were either discarded or used for downstream applications. It is important that cells never reach confluency during routine cultivation, since this negatively affects the cells' potential to differentiate.

#### 2.1.2 MYOTUBE DIFFERENTIATION

Cells were seeded at a density of  $8 \times 10^4$  cells/cm<sup>2</sup> and incubated in differentiation medium (high glucose DMEM supplemented with 1 % Penicillin/Streptomycin (Sigma-Aldrich) and 2 % horse serum (Thermo Fisher) at 37 °C and 5 % CO<sub>2</sub>. Differentiation medium was replaced every 2 days. Typically, after 5 to 6 days contracting myotubes have formed.

#### 2.1.3 POLY-L-LYSINE COATING OF COVERSGLIPS

15 x 15 mm square coverslips (R. Langenbrinck or Carl Roth) were thoroughly cleaned with 80 % EtOH and coated with poly-L-lysine. Therefore, a drop of 1 mg/ml poly-L-lysine (Sigma-Aldrich) was placed on parafilm and a coverslip was positioned on top. The coverslips were incubated with poly-L-lysine for 15 min. Then, coverslips were carefully picked up with forceps and plunged into a beaker containing ddH<sub>2</sub>O to wash off residual poly-L-lysine. Coverslips were placed into 12-well plates and air-dried in the fume hood. To ensure sterility, 12-well plates with coverslips were then placed in a tissue culture hood under UV light for 1 h without lids.

### 2.1.4 TRANSCRIPTION INHIBITOR TREATMENT

*Note: As transcription inhibitors are highly toxic, it is mandatory to wear protective equipment (gloves and eye protection). Medium supplemented with transcription inhibitors needs to be collected separately and disposed appropriately.*

#### **Actinomycin D**

Differentiation medium was supplemented with 10 µg/ml actinomycin D (Sigma-Aldrich) and cells were incubated at 37 °C and 5 % CO<sub>2</sub> for 12 h before fixation.

#### **α-Amanitin**

Differentiation medium was supplemented with 25 µg/ml α-amanitin (Sigma-Aldrich) and cells were incubated at 37 °C and 5 % CO<sub>2</sub> for 24 h before fixation.

#### **5,6-Dichloro-1-β-D-ribofuranosylbenzimidazol (DRB)**

Differentiation medium was supplemented with 100 µM DRB (Sigma-Aldrich) and cells were incubated for different timepoints (2 min, 25 min, 50 min, 75 min, 3 h and 6 h) at 37 °C and 5 % CO<sub>2</sub> before fixation. In a second experiment, cells were treated with 100 µM DRB for 6 h, washed with PBS twice and incubated in differentiation medium without DRB for different timepoints (25 min, 50 min, 75 min and 3 h) before fixation.



## 2.2 TISSUE PREPARATION

### 2.2.1 ISOLATION AND FIXATION OF MOUSE TISSUES

CD1 exbreeder mice (Charles River Laboratories) were sacrificed by cervical dislocation after anesthetizing with IsoFlo (Isofluran, Abbott) by Dr. Irina Solovei.

For isolation of the thyroid gland, a vertical anterior neck incision was made. After removing the parotid glands and surrounding muscle tissue, the thyroid lobes were visible below the thyroid cartilage on each side of the trachea. A piece of trachea (5 to 7 mm) with the attached thyroid lobes was excised and placed into PBS in a small plastic dish. Under a binocular, the thyroid tissue was carefully cleaned from connective and fat tissue, muscles as well as the parathyroid glands (dense, slightly whitish tissue) using fine forceps and fine scissors (both Fine Science Tools).

For isolation of colon, a vertical incision in the lower abdomen was made and extended horizontally in both directions. The abdominal musculature was carefully removed, a part of the distal colon was excised and placed into PBS in a small plastic dish.

For isolation of bladder, mice were injected with Narcoren (Boehringer Ingelheim, conc. 5 µl/g body weight) by Hilde Wohlfrom (Biocenter, LMU München, Munich, Germany) in order to relax smooth muscles before sacrificing. The abdominal musculature was carefully removed, the bladder was excised and placed into PBS in a small plastic dish.

All tissues were washed once with PBS and then fixed in 4 % paraformaldehyde (Carl Roth) solution in PBS overnight.

Sections of *in vitro* grown thyroid follicles as well as grafted thyroid tissue were kindly provided by Dr. Andrea Schiavo and Dr. Sabine Costagliola (Université libre de bruxelles, Brussels, Belgium) (Antonica et al., 2012).

Fixed thyroids of hypo- and hyperthyroid mice and of mice in different developmental stages were kindly provided by Prof. Dr. Heike Heuer (Universitätsklinikum Essen, Klinik für Endokrinologie und Stoffwechselerkrankungen, Essen, Germany). Hypo- and hyperthyroid mice were obtained as previously described (Groba et al., 2013; Shibusawa et al., 2003; Trajkovic-Arsic et al., 2010) (see also Results section 3.7.4).

### 2.2.2 THYROID TISSUE OF OTHER VERTEBRATES

*Danio rerio*, expressing nuclear eGFP under the TG promoter, were kindly provided by Dr. Sabine Costagliola (Université libre de bruxelles, Brussels, Belgium) (Opitz et al., 2013).

Dissected thyroid glands of *Xenopus tropicalis* females were kindly provided by Dr. Jean-Baptiste Fini (Centre national de la recherche scientifique CNRS, Paris, France).

Chicken thyroid glands were kindly provided by Prof. Dr. med. vet. Bernd Kaspers (Institut für Tierphysiologie, LMU München, Munich, Germany).

Human thyroid gland samples were kindly provided by Prof. Dr. Heike Heuer (Universitätsklinikum Essen, Klinik für Endokrinologie und Stoffwechselerkrankungen, Essen, Germany).

### 2.2.3 EMBEDDING AND CRYOSECTIONING OF TISSUE

After fixation, mouse tissues were washed in PBS and incubated in increasing concentrations of sucrose: 10 % for 1 h, 20 % for 1 h and 30 % for 1 to 12 h. For storage, tissues were frozen at  $-20^{\circ}\text{C}$  in 30 % sucrose solution. For embedding, tissues were incubated for 10 min in Tissue-Tek O.C.T. compound freezing medium (Sakura). A drop of freezing medium was placed on the bottom of a small flexible silicone rubber mold (Pelco) and the tissue was placed on top and oriented in the desirable angle. The mold was filled up with freezing medium to cover the whole tissue and the molds were placed into a container with dry ice. Once the medium with tissue was frozen, blocks were stored in boxes at  $-80^{\circ}\text{C}$ .

Frozen mouse tissues were cut into  $16\text{ }\mu\text{m}$  sections using a cryostat (Leica CM3050S). Chamber temperature was set to  $-20^{\circ}\text{C}$  and object temperature was set to  $-18^{\circ}\text{C}$ . Immediately after cutting, the section adheres to the knife and a microscopic slide with increased adhesion properties (Superfrost Plus, Thermo Scientific) was brought close to the section so that the section would stick to it. Slides with cryosections were kept in plastic boxes in the cryostat chamber while sectioning. For long term storage, the boxes were kept at  $-80^{\circ}\text{C}$ .

## 2.3 TRANSCRIPTION STIMULATION VIA DCAS9-VPR

To activate transcription, we employed a Cas9-based approach described by Chavez and colleagues (Chavez et al., 2015). A nuclease dead Cas9 (dCas9) fused to a tripartite transcriptional activator termed VPR (consisting of VP64, p65 and Rta) and targeted to the upstream promoter region of a gene of interest via a set of different gRNAs. Therefore, we stably integrated a cassette to inducibly express dCas9-VPR into Pmi28 mouse myoblasts via the PiggyBac transposon system. dCas9-VPR expression was induced by addition of 1 µg/ml doxycycline (Sigma-Aldrich).

### 2.3.1 gRNA EXPRESSION PLASMID GENERATION

To monitor gRNA transfection efficiency, a plasmid expressing GFP-H2A additionally to the respective gRNA (U6-gRNA-GFP-H2A) was generated. CMV-GFP-H2A was inserted into the pEX-A-U6-gRNA expression plasmid via Gibson assembly. pEX-A-U6-gRNA was synthesized at Eurofins MWG Operon according to (Mali et al., 2013). Gibson assembly was performed using pEX-A-U6-gRNA linearized by SacI and GFP-H2A generated by PCR using the following primers:

Fwd (5' → 3'): atatgggtaccgagctTAGTTATTAATAGTAATCAATTACGGG

Rev (5' → 3'): ttgcggccgcgagctTGATCAGTTATCTAGATCCGG

gRNA vectors were generated by amplifying U6-gRNA-GFP-H2A with forward and reverse primers which introduced the protospacer sequence for the respective promoter target region. Primer sequences for introducing protospacer sequences are listed in the Appendix, Table S1.

### 2.3.2 STABLE DCAS9-VPR EXPRESSION

Pmi28 mouse myoblasts were transfected with PB-TRE-dCas9-VPR (purchased from Addgene, ID 63800, laboratory internal identifier pc4538) and PiggyBac transposase (purchased from System Biosciences, #PB200PA-1, laboratory internal identifier pc3015) expression plasmids at a ratio of 3:1 with Lipofectamine 3000 (Invitrogen) according to manufacturer's instructions. Clones that stably integrated dCas9-VPR were selected with medium containing 50 µg/ml hygromycin B (Sigma-Aldrich) over a time course of 2 weeks. Single cells were sorted into 96 well plates by fluorescence activated cell sorting (FACS) (BD Biosciences Aria II Sorter) and cultured for one week in presence of hygromycin B. dCas9-VPR expression was induced with doxycyclin at a final concentration of 1 µg/ml for 50 h. Individual clones were tested for dCas9-expression by western blot using a Cas9 antibody (Clontech, dilution 1:1000). To verify that dCas9-VPR could be correctly targeted to a desired genomic region, stable Pmi28 clones were transfected with gRNA expression plasmids containing the protospacer sequences for major and minor satellites and telomere repeats, respectively, as described previously (Anton et al., 2014). Localization of dCas9-VPR was assessed

via immunofluorescence staining (see Materials and Methods Section 2.4.9) with Cas9 antibody (Clontech, dilution 1:150).

### 2.3.3 TRANSFECTION OF gRNA PLASMIDS

For gRNA transfection,  $3 \times 10^5$  cells were seeded per 6 well on coverslips and transfected with 2.5  $\mu\text{g}$  of a mix of 6 different gRNA plasmids per target gene (see Appendix, Table S1) at equimolar amounts with Lipofectamine 3000 (Invitrogen) according to manufacturer's instructions. Medium was removed after 12 h and replaced by medium containing 1  $\mu\text{g}/\text{ml}$  doxycyclin to induce dCas9-VPR expression.

### 2.3.4 ANALYSIS

After 50 h of doxycycline treatment, cells were fixed and subjected to ImmunoFISH as described in section 2.4.9. A mouse anti-GFP antibody (Roche, dilution 1:400) together with a donkey anti-mouse Alexa 488 antibody (Invitrogen, dilution 1:500) were used to identify cells that are transfected with gRNAs. The FISH probe was generated by nick translation (see Materials and Methods section 2.4.3) from BAC clone RP23-310F9, encompassing the *Ttn* gene.

The relative expression level of *Ttn* and *Neb* was assessed by quantitative real time PCR (see Materials and Methods section 2.5.2) using the housekeeping gene glyceraldehyde 3-phosphate dehydrogenase (*Gapdh*) as reference. qPCR was performed with one biological replicate in technical triplicates.

## 2.4 FLUORESCENCE IN SITU HYBRIDIZATION (FISH)

### 2.4.1 LABELING OF dUTPs

Labeling of dUTPs was performed as described previously (Cremer et al., 2008) with modifications. 1 mg of aminoallyl-dUTP (AA-dUTP, Sigma-Aldrich) was dissolved in 96  $\mu$ l 0.2 M bicarbonate buffer ( $\text{NaHCO}_3$ ), yielding a final concentration of 20 mM. Fluorochromes and haptens were diluted in DMSO according to Table 2.1.

**TABLE 2.1: Fluorochrome/hapten dilutions in DMSO**

Fluorochrome/Hapten	Manufacturer	Delivered quantity	DMSO added	Final conc.
Biotin succinimidyl ester (BIO)	Invitrogen	100 mg	4,401 $\mu$ l	40 mM
Digoxigenin succinimidyl ester (DIG)	Invitrogen	5 mg	213 $\mu$ l	40 mM
FITC succinimidyl ester	Invitrogen	10 mg	417 $\mu$ l	40 mM
Cy3 mono NHS ester	Amersham Biosciences	1 mg	66 $\mu$ l	20 mM
TAMRA succinimidyl ester	Invitrogen	10 mg	1,560 $\mu$ l	10 mM
Texas Red succinimidyl ester	Invitrogen	5 mg	612 $\mu$ l	10 mM
Cy5 mono NHS ester	Amersham Biosciences	1 mg	62 $\mu$ l	20 mM

To conjugate fluorophores/haptens to dUTPs, a reaction was set up according to Table 2.2 and incubated at RT for 3 to 4 h.

**TABLE 2.2: Conjugation reaction set-up**

Ingredient	Volume
20 mM AA-dUTP	10 $\mu$ l
H <sub>2</sub> O	10 $\mu$ l
0.2 M bicarbonate buffer	10 $\mu$ l
40 mM <b>FITC</b> <u>or</u> 20 mM <b>Cy3</b> <u>or</u> 10 mM <b>TAMRA</b> <u>or</u> 10 mM <b>TexasRed</b> <u>or</u> 20 mM <b>Cy5</b>	10 $\mu$ l (for Texas Red use 20 $\mu$ l)

2  $\mu$ l 2 M glycine (pH 8.0) and 4  $\mu$ l 1 M Tris-HCl (pH 7.75) were added to stop the reaction and stabilize the nucleotides. Then, the volume was adjusted to 200  $\mu$ l with ddH<sub>2</sub>O, yielding 1 mM solutions of the respective conjugated dUTP. Labeled nucleotides were stored at  $-20^\circ\text{C}$ .

## 2.4.2 PREPARATION OF dNTP MIX FOR NICK TRANSLATION AND DOP-PCR

For convenience, a mix of unlabeled dNTPs and labeled dUTPs was prepared in advance for later use in nick translations. First, a 1 mM solution of dATP/dGTP/dCTP was prepared by adding 4 µl of each dNTP (100 mM, Roche) to 388 µl ddH<sub>2</sub>O. Second, a 1 mM solution of dTTP was prepared by adding 4 µl of dTTP (100 mM, Roche) to 396 µl ddH<sub>2</sub>O. To prepare the dNTP mix (1 mM), combine 50 µl 1 mM dATP/dGTP/dCTP, 25 µl 1 mM dTTP and 25 µl 1 mM fluorophore/hapten-labeled dUTP.

## 2.4.3 NICK TRANSLATION (NT)

DNA was labeled with fluorophores or haptens by nick translation (NT). In this reaction, lowly concentrated DNase I induces single strand breaks into the DNA. DNA Pol I fills these nicks by elongating the 3' hydroxy terminus while it removes nucleotides with its 5' to 3' exonuclease activity. Thereby, labeled nucleotides are incorporated. Typically, 1 µg of DNA was labeled in a reaction volume of 50 µl (Table 2.3).

**TABLE 2.3 : Nick translation set-up**

Probe DNA*	NT buffer*	0.1 M β-ME	dNTP Mix	H <sub>2</sub> O	DNase I (1:250) (Roche)	DNA Pol I (Thermo Scientific)
1 µg	5 µl	5 µl	5 µl	(33-x) µl	1 µl	1 µl

\* See notes to this section

The reaction was incubated at 15 °C in a waterbath for 2 h. Labeled DNA was column purified (NucleoSpin Gel and PCR Clean-up Kit, Macherey-Nagel) and eluted in 25 µl H<sub>2</sub>O (half the reaction volume). The volume was adjusted by adding 25 µl (half the reaction volume) blue stop mix (0.1 % bromophenol blue, 0.5 % dextran blue, 0.1 M NaCl, 20 mM Tris-HCl pH 7.5).

Notes :

- The reaction is scalable up to 100 µl.
- Typically, 4 µl of GenomiPhi amplified BAC DNA were used for NT in 50 µl. However, depending on the BAC, volumes might need to be adjusted.
- DNase I dilution can vary; one should figure out the exact dilution for every new batch of DNase I.
- 10 x DNA Pol I buffer is used as NT buffer. Alternatively, 10 x NT buffer (0.5 M Tris-HCl (pH 7.5) supplemented with 50 mM MgCl<sub>2</sub> and 0.05 % BSA) can be prepared.

#### 2.4.4 FISH PROBE DESIGN AND PREPARATION

##### BAC DNA

BAC clones encompassing the desired genomic region were selected using the UCSC genome browser (<https://genome.ucsc.edu/>) and purchased from BACPAC Resources (Oakland children's hospital) as agar stabs (<https://bacpacresources.org/>). Bacteria were grown in LB medium containing 12.5 µg/ml chloramphenicol and BACs were purified via standard alkaline lysis (2 ml of bacterial culture) or the NucleoBond Xtra Midi Kit (Macherey-Nagel) (200 ml of bacterial culture) and amplified using the GenomiPhi Kit (GE Healthcare) according to the manufacturer's instructions. BAC probes were labeled with the desired fluorophores or haptens by NT (see Materials and Methods section 2.4.3). BAC clones used in this study are listed in the Appendix, Table S2.

##### Tg exon probes

Probes to specifically label exons 2 to 12 and 33 to 47 of *Tg* were generated by standard polymerase chain reaction (PCR) using a cDNA clone (transOMIC technologies, BC111467) as template. Primer sequences are indicated in the Appendix, Table S3.

2 µg of PCR amplified DNA were labeled with fluorophores by NT (see Material and Methods section 2.4.3) in 100 µl reactions. Incubation time at 15 °C was reduced to 1.5 h due to the small size of the DNA fragments (3,045 bp and 2,143 bp for Ex 2 to 12 and Ex 33 to 47, respectively) in order to avoid excessive digestion by DNase I.

##### Tg start and intron 41 probes

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

Primers to amplify fragments for FISH probes were designed with the HD FISH probe generator platform accessible via <http://www.hdfish.eu/> (Bienko et al., 2013). Probe generation was conducted as described previously (Bienko et al., 2013) in 96 well plates (LightCycler 480 Multiwell, Roche) using a LightCycler 480 Instrument II (Roche) and KAPA SYBR FAST (Roche) (Tables 2.4 and 2.5).

Primers were ordered from Eurofins Genomics (Ebersberg, Germany) in a 96 well plate format in mixed plates, with each well containing both the forward and respective reverse primer at a concentration of 5 µM. Primer sequences for amplification of probes specific for the start of *Tg* and the first and second half of *Tg* intron 41 are displayed in the Appendix Tables S4, S5 and S6, respectively.

**TABLE 2.4: Reaction set-up for probe amplification**

	For 1 well	For 96 well plate (x100)
<b>KAPA SYBR FAST 2x</b>	25 $\mu$ l	2,500 $\mu$ l
<b>gDNA 50 ng/<math>\mu</math>l in TE</b>	1 $\mu$ l	100 $\mu$ l
<b>nuclease-free water</b>	20 $\mu$ l	2,000 $\mu$ l
	46 $\mu$ l	

4  $\mu$ l of the 5  $\mu$ M forward and reverse primer dilution was added to the corresponding well.

**TABLE 2.5: LightCycler 480 II settings for probe amplification**

	Target [°C]	Acquisition mode	Hold [hh:mm:ss]	Ramp rate [°C/s]	Acquisitions [per °C]
Pre-Incubation	95	none	00:05:00	4.4	-
Amplification	95	none	00:00:10	4.4	-
	55	none	00:00:15	2.2	-
	72	single	00:00:30	4.4	-
Melting Curve	95	none	00:00:05	4.4	-
	65	none	00:01:00	2.2	-
	97	continuous	-	-	10
Cooling	40	none	00:00:10	1.5	-

After amplification in 96 well plates, fragments corresponding to one probe were mixed in a 15 ml tube and EtOH precipitated.

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



### Chromosome paints

Whole chromosome paints were a kind gift from Dr. Nigel Carter and Dr. Johannes Wienberg (Cambridge University, Department of Pathology, Cambridge, United Kingdom) and amplified as well as fluorophore/hapten labeled with biotin or digoxigenin via DOP-PCR with a partially degenerate oligonucleotide primer (6MW primer: 5'- CCG ACT CGA GNN NNN NAT GTG G -3', Eurogentec) according to tables 2.6 and 2.7.

**TABLE 2.6: Set-up for labeling DOP-PCR**

	Start concentration	Volume	Final concentration
PCR buffer (PerkinElmer)	10 x	5 µl	1 x
MgCl <sub>2</sub> (PerkinElmer)	25 mM	4 µl	2 mM
6MW primer	20 µM	5 µl	2 µM
dATP/dGTP/dCTP-Mix	2 mM	2.5 µl	100 µM
dTTP	1 mM	4 µl	80 µM
Label-dUTP	1 mM	1 µl	20 µM
H <sub>2</sub> O	-	26.5 µl	-
		48 µl*	

\*To this mixture, 1 µl DNA and 1 µl Taq polymerase were added (total volume 50 µl).

**TABLE 2.7: Cycling conditions for labeling DOP-PCR**

	Temperature	Time	20 cycles
Initial denaturation	94°C	3 min	
Denaturation	94°C	1 min	
Annealing	56°C	1 min	
Extension	72°C	30 sec	
Final extension	72°C	5 min	
	4°C	hold	

### 2.4.5 PREPARATION OF HYBRIDIZATION MIXTURE

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

#### Hybridization mixture for 2D FISH with BAC derived probe (6 µl: 3 µl FA + 3 µl MM)

Chromosome paint - BIO	15 µl
BAC - DIG	25 µl
Cot-1 DNA (Thermo Fisher)	20 µl
Salmon sperm DNA (Thermo Fisher)	2 µl
Pre-cooled (−20 °C) 100 % ethanol	300 µl

#### Hybridization mixture for 3D FISH with BAC derived probe (12 µl: 6 µl FA + 6 µl MM)

BAC - fluorophore/hapten	100 µl
Cot-1 DNA (Thermo Fisher)*	25 µl
Salmon sperm DNA (Thermo Fisher)*	3 µl
Pre-cooled (−20 °C) 100 % ethanol*	500 µl

\*If more BACs were to be combined in the same probe, the volume of Cot-1, salmon sperm and EtOH were upscaled accordingly.

#### Hybridization mixture for 3D FISH with PCR derived probes (5 µl: 2.5 µl FA + 2.5 µl MM)

PCR probe - fluorophore	10 µl
Salmon sperm DNA (Thermo Fisher)*	10 µl
Pre-cooled (−20 °C) 100 % ethanol*	250 µl

\*If more than one PCR-derived probe were to be combined in the same FISH probe, the volume of salmon sperm and EtOH were upscaled accordingly.

The mixture was incubated for > 2 h (or overnight) at −20 °C and spun down at 14,000 rpm for 30 min at 4 °C in a tabletop centrifuge. To remove residual traces of EtOH, pellets were dried in a SpeedVac (Vacuum concentrator, Bachofer) with heating for approx. 5 min or until no liquid was visible. The pellet was dissolved in 100 % ultrapure deionized formamid (FA, VWR Amresco) by alternating incubation in the waterbath at 50 °C and thorough vortexing. The addition of FA to the hybridization mixture affects the thermodynamic stability of the DNA double helix and enables to carry out the hybridization at lower temperatures. Once the pellet had dissolved, an equal volume of the so-called master mix (MM; 20 % dextran sulfate in 2 x SSC) was added to the probe and mixed by vortexing. Dextran sulfate acts as a volume exclusion agent and therefore increases the local concentration of probe DNA enhancing the hybridization rate. Hybridization mixtures were kept at −20 °C for long term storage.

#### 2.4.6 2D FISH

2D FISH experiments were carried out on metaphase spreads of mouse embryonic fibroblasts (MEFs) to verify the correct position of the selected BACs on the respective chromosome. BACs were labeled with digoxigenin via NT and chromosome paints were labeled with biotin via DOP-PCR.

Slides with metaphase spreads were withdrawn from the freezer approx. 30 min prior to usage to enable temperature acclimatization. The regions containing metaphase spreads were marked on the backside of the microscopic slide using a diamond cutter and hybridization mixture was directly placed on top of the spreads. The sample area was covered with a 15 x 15 mm coverslip and the edges were sealed with rubber cement (Fixogum) to prevent drying during hybridization. Once the rubber cement was completely dry, slides were placed on a heatblock set to 78 °C for 3 min to simultaneously denature target DNA and probe DNA.

Hybridization was carried out at 37 °C in a chamber in the waterbath overnight. Post hybridization washings were carried out using 2 x SSC at 37 °C (3 x 10 min) and 0.1 x SSC at 60 °C (1 x 5 min). Detection was carried out using a biotin binder (streptavidin-Alexa 555, Molecular Probes, dilution 1:200) and an antibody against digoxigenin (sheep  $\alpha$  DIG FITC, Roche, dilution 1:200). Antibodies were diluted in blocking solution (4 % BSA/4 x SSCT). Antibody incubation was carried out for 45 to 60 min. Washing was performed using 4 x SSCT (2 x 10 min). DNA was counterstained with 2  $\mu$ g/ml DAPI (Sigma-Aldrich) for 10 min in a dark chamber. Preparations were washed once with PBS and once with distilled H<sub>2</sub>O and air dried in the darkness. A drop of Vectashield (Vector Laboratories) was placed on the sample area and the area was mounted under a coverslip. Excess Vectashield was removed and the edges of the coverslip were sealed with nail polish.

#### 2.4.7 3D FISH ON TISSUE CRYOSECTIONS

FISH on cryosections was conducted as previously described (Solovei et al., 2007) with modifications. In general, incubation and washing steps were conducted by placing the slides with the section into Coplin jars with the corresponding buffers. When a probe was mounted, the slides were kept in darkness to avoid photobleaching.

Cryosections were dried at RT for 20 to 30 min. Slides were then placed into 10 mM sodium citrate buffer (pH 6.0) for 3 min to re-hydrate the sections before they were transferred to a Coplin jar containing pre-heated 80 °C sodium citrate buffer and incubated for 20 min to remove fixation-derived antigen masking (antigen retrieval). For DNA FISH, sections were treated with RNase to ensure hybridization only to DNA. Therefore, slides were briefly washed in PBS, sections were covered with a drop of 50  $\mu$ g/ml RNase and a coverslip was placed on top. Incubation was carried out in a metal chamber in the water bath at 37°C for 1h followed by washing in PBS (3 x 30 min). For both DNA and RNA FISH, slides were then briefly equilibrated in 2 x SSC and incubated in 50 % formamid/2 x SSC for 30 min to 1 h. Slides were carefully dried from

excess liquid using a tissue paper and sections were covered by special glass chambers made from coverslips in order to prevent tissue squeezing (Solovei et al., 2007).

For DNA FISH, a probe was loaded under the chamber, the edges of the chamber were sealed with rubber cement and the slides were put into darkness until the rubber cement had fully dried out. Slides were then placed into a metal chamber in the water bath at 37 °C in order to equilibrate the section in the hybridization mixture and to ensure probe penetration. Denaturation of both probe and sample DNA was carried out simultaneously on a hot block at 80 °C for 3 min.

For RNA FISH, only the probe DNA, but not the sample DNA was denatured, so that the then single stranded probe only hybridizes to single stranded sample RNA and not to double stranded sample DNA. Therefore, the probe DNA was denatured in the water bath at 95 °C for 5 min, before it was loaded under the glass chamber and sample DNA denaturation was omitted.

Hybridization was carried out in the water bath at 37 °C for 2 days. Post hybridization washings were carried out using 2 x SSC at 37 °C (3 x 30 min) and 0.1 x SSC at 60 °C (1 x 7 min). For some experiments, depending on the probe label, detection was necessary. Antibodies used were the same as described for 2D FISH. Antibody incubation times were prolonged to 12 h to ensure proper antibody penetration into the tissue. Washings were performed using 4 x SSCT (3 x 30 min).

DNA was counterstained with 2 µg/ml DAPI (Sigma-Aldrich) for 1 h at 37 °C in the water bath. Sections were washed once in PBS, a drop of Vectashield (Vector Laboratories) was placed directly on the section and the area was mounted under a coverslip. Excess Vectashield was carefully removed in order to keep fragile sections intact and the edges of the coverslip were sealed with nail polish.

### 2.4.8 3D FISH ON ADHERENTLY GROWING CELLS

FISH on adherently growing cells was conducted as previously described (Solovei and Cremer, 2010) with modifications. Washing as well as incubation steps for cells were carried out in 6 well or 12 well plates depending on the coverslip size used.

Cells were grown on coverslips (R. Langenbrinck or Carl Roth) pre-coated with 1 µg/ml poly-L-lysine (Sigma-Aldrich). Coverslips were washed twice with pre-warmed (37 °C) cell culture grade PBS (Sigma-Aldrich) and fixed in 4 % PFA (Merck) for 10 min at RT. During the last minute of fixation, a drop of PBS/0.5 % Triton X-100 was added to prevent cell drying in subsequent steps. Cells were washed 3x in PBS/0.01 % Triton X-100 for 5 min, permeabilized with PBS/0.5 % Triton X-100 for 10 min and then equilibrated in 20 % glycerol/PBS for 60 min. Cells were frozen by submerging the coverslip directly into liquid nitrogen for approx. 20 sec, then, coverslips were gradually thawed at RT, before they were frozen again. This freezing/thawing cycle was repeated four times. Then, cells were washed three times with PBS/0.01 % Triton X-100 for 10 min. For DNA FISH, cells were first treated with 0.1 N HCl for 5 min sharp in order to

depurinate DNA and remove proteins, improving access of the probe to the DNA. Then, cells were treated with RNase to ensure hybridization only to DNA. Therefore, coverslips were briefly washed in PBS, and placed on drops of 50 µg/ml RNase on a piece of parafilm with attached cells facing downwards. Incubation was carried out in a metal chamber in the water bath at 37°C for 1h followed by washing in PBS (3 x 10 min). For both DNA and RNA FISH, cells were rinsed in 2 x SSC and equilibrated in 50 % FA/2 x SSC.

For DNA FISH, a probe was loaded on a microscopic slide, the coverslip with the cells facing the slide was carefully placed on top to avoid bubble formation and the edges of the coverslip were sealed with rubber cement (Fixogum). The slides were put into darkness until the rubber cement had fully dried out. Denaturation of both probe and sample DNA was carried out simultaneously by placing the slide on a hot block at 78 °C for 3 min.

For RNA FISH, only the probe DNA, but not the sample DNA was denatured, as described for FISH on sections, before it was loaded on the microscopic slide and the coverslip with cells was placed on top. Sample DNA denaturation was omitted.

Hybridization was carried out in the water bath at 37 °C for 2 days. Post hybridization washings were carried out using 2 x SSC at 37 °C (3 x 30 min) and 0.1 x SSC at 60 °C (1 x 7 min). For some experiments, depending on the probe label, detection was necessary. Antibodies used were the same as described for 2D FISH. Washings were performed using 4 x SSC<sup>T</sup> (3 x 10 min).

DNA was counterstained with 2 µg/ml DAPI (Sigma-Aldrich) for 10 min in a dark chamber. Coverslips were washed once in PBS, a drop of Vectashield (Vector Laboratories) was placed on a microscopic slide and the coverslip was carefully mounted on top avoiding bubble formation. Excess Vectashield was very carefully removed from the edges of the coverslip in order to not move the coverslip and the edges of the coverslip were sealed with nail polish.

## 2.4.9 IMMUNOFISH

### ImmunofISH on cryosections

To visualize the co-localization of RNA Pol II or nuclear speckles (SC35) and the *Tg* transcription loop, immunostaining combined with 3D FISH was conducted on mouse (for RNA Pol II) or human (for SC35) thyroid cryosections. The general 3D FISH protocol on cryosections (see Material and Methods section 2.4.7) was followed, instead of DAPI staining directly after the post-hybridization washings, slides were briefly equilibrated in PBS. Permeabilization was done in 0.1 % Triton X-100/PBS for 30 min. The sections were then incubated with primary and secondary antibodies diluted in blocking solution (1 % BSA/0.1 % Triton X-100/0.1 % saponin) each overnight at RT under a glass chamber in a humidified chamber in the dark. DAPI (2 µg/ml) was added directly to the secondary antibody.

### Primary antibodies:

- mouse  $\alpha$ -RNA Pol II S2P, Pc26, dilution 1:100 (kind gift from Prof. Dr. Hiroshi Kimura, Tokyo Institute of Technology, Tokyo, Japan)
- mouse  $\alpha$ -SC35 (Sigma-Aldrich), dilution 1:50

Secondary antibody: donkey  $\alpha$ -mouse Alexa 488 (Invitrogen), dilution 1:250

Washings in between antibodies and after the secondary antibody were performed with 0.01 % Triton X-100/PBS at 37 °C three times for 30 min. Sections were briefly washed with PBS and then mounted in Vectashield.

### ImmunoFISH on cells

ImmunoFISH on adherently growing cells was performed as previously described (Solovei and Cremer, 2010) with modifications. To visualize the *Tm* gene together with H2A-GFP indicating successful transfection of gRNAs targeting *Tm* promoter regions, immunostaining combined with 3D FISH was conducted on mouse Pmi28 myoblasts. Coverslips were washed in PBS prewarmed to 37 °C twice, before cells were fixed in 4 % PFA/PBS for 10 min at RT. Cells were permeabilized with 0.5 % Triton X-100/PBS for 10 min at RT and incubated with primary and secondary antibodies diluted in blocking solution (PBS/4 % BSA/0.01 % Tween 20) for 1 h each.

- Primary antibody: mouse  $\alpha$ -GFP (Roche), dilution 1:400
- Secondary antibody: donkey  $\alpha$ -mouse Alexa 488 (Invitrogen), dilution 1:500

Washings in between antibodies and after the secondary antibody were performed with PBS/0.1 % Tween 20. Cells were post-fixed in 2 % PFA/PBS for 10 min and then incubated in 20 % glycerol for 1 h. From here on, the standard protocol for FISH on adherently growing cells was followed.

## 2.5 GENE EXPRESSION ANALYSIS

### 2.5.1 RNA ISOLATION

Total RNA from cells or tissues was isolated using the NucleoSpin RNA Kit (Macherey-Nagel) according to the manufacturer's instructions. To enable proper lysis, tissues were disrupted directly in 350 µl lysis buffer RA1 using a Tissue Homogenizer (Bullet Blender 24, Next Advance) with zirconium oxide beads with a diameter of 0.5 mm for 5 to 10 minutes until a homogeneous solution had formed, prior to applying the sample to the clearing column (violet ring). RNA integrity was checked by separating RNA fragments on a 1 % agarose (Carl Roth) gel. Only samples with a 28S:18S rRNA ratio of  $\approx 2:1$  and without genomic contamination and RNA degradation were used for downstream applications.

### 2.5.2 cDNA PREPARATION AND QUANTITATIVE REAL TIME PCR (qPCR)

1 µg of total RNA was reverse transcribed using the High capacity cDNA reverse transcription Kit (Applied Biosystems) according to the manufacturer's instructions. A sample without reverse transcriptase served as control.

Quantitative PCR (qPCR) was performed in technical and biological triplicates in 10 µl reactions on a LightCycler 480 Instrument II (Roche) using LightCycler 480 SYBR Green Master Mix (Roche) according to the manufacturer's instructions. Primers (Appendix, Table S7) were used at a final concentration of 250 nM. Primer specificity was ensured by melting curve analysis and PCR efficiency was determined using a standard curve of serially diluted cDNA. Primers were designed specifically covering exon-exon junctions via NCBI primer design tool accessible at <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>.

Typically, 1 ng of cDNA were used per qPCR reaction. Due to its high transcript abundance, 50 pg of cDNA were used per reaction to determine the *Tg* transcription levels in different developmental stages as well as in hypo- and hyperthyroid mice. Amplification was run according to Table 2.8.

**TABLE 2.8: Settings for cDNA amplification**

	Target [°C]	Acquisition mode	Hold [hh:mm:ss]	Ramp rate [°C/s]	Acquisitions [per °C]
Denaturation	95	none	00:05:00	4.4	-
Amplification	95	none	00:00:10	4.4	-
	60	none	00:00:10	2.2	-
	72	single	00:00:20	4.4	-
Melting Curve	95	none	00:00:10	4.4	-
	65	none	00:00:30	2.2	-
	95	continuous	-	-	5
Cooling	40	none	00:00:30	2.2	-

CT (cycle threshold) values were normalized to the transcript level of thyroid specific transcription factor paired box gene 8 (*Pax8*) for *Tg* or glyceraldehyde 3-phosphate dehydrogenase (*Gapdh*) for *Ttn* and *Neb* and further analyzed based on the  $2^{-\Delta\Delta CT}$  method.

### 2.5.3 RNA SEQUENCING (RNA-SEQ)

#### Sample and library preparation

Total RNA was isolated as described in section 2.5.1. Sequencing libraries for RNA-seq were prepared by Christopher Mulholland (BioSysM, LMU München, Munich, Germany) as follows: digital gene expression libraries for RNA-seq were produced using a modified version of single-cell RNA barcoding sequencing (SCR-seq) optimized to accommodate bulk cells (Bagnoli et al., 2018; Ziegenhain et al., 2017) in which a total of 70 ng of input RNA was used for the reverse-transcription of individual samples.

#### Sequencing

Sequencing was performed at the laboratory for functional genome analysis (LAFUGA, Gene Center, LMU München, Munich, Germany). RNA-seq libraries were sequenced on an Illumina HiSeq 1500.

#### Analysis

Bioinformatical analyses were performed by Dr. Sebastian Bultmann (BioSysM, LMU München, Munich, Germany). RNA-seq libraries were processed and mapped to the mouse genome (mm10) using the zUMIs pipeline (Parekh et al., 2018). UMI count tables were filtered for low counts using HTSFilter (Rau et al., 2013).

GENCODE gene annotation version vM11 (mouse) was downloaded from <https://www.encodegenes.org/>. Using R (<https://www.R-project.org/>), mitochondrial genes were excluded and only protein-coding genes with a known gene status were selected to ensure proper annotation and used for further evaluation. Gene lengths were calculated from GENCODE annotated start and end positions of each gene.



#### 2.5.4 NANOPORE SEQUENCING (NANOPORE-SEQ)

##### **Sample preparation**

Mouse thyroid glands were excised as described in section 2.2.1. For each sample, the glands of five mice were used. The glands were cut into small pieces (< 1 mm) with fine scissors under a binocular in RNase free PBS and directly transferred to TRIzol (Thermo Fisher). Samples were stored at  $-80^{\circ}\text{C}$ . Total RNA was isolated using the TRIzol method at the Laboratory for Functional Genome Analysis (LAFUGA, Gene Center, LMU München, Munich, Germany).

##### **Library preparation and sequencing**

Library preparation for direct RNA sequencing and sequencing itself were performed at the Laboratory for Functional Genome Analysis (LAFUGA, Gene Center, LMU München, Munich, Germany). Due to the low concentration of starting material, no polyA enrichment was performed, but total RNA was used for library preparation directly.

##### **Analysis**

Nanopore sequencing alignment and analysis were conducted by Dr. Sebastian Bultmann (BioSysM, LMU München, Munich, Germany). Reads were mapped against the mouse genome (mm10) using minimap2 with standard parameters. Coverage of introns and exons was calculated using Bedtools (Quinlan et al., 2010) and custom R scripts.

## 2.6 NATIVE CHROMATIN IMMUNOPRECIPITATION FOLLOWED BY SEQUENCING (NCHIP-SEQ)

Native ChIP-seq (nChIP-seq) experiments were performed with help and advice from Stephanie Link (Biomedical Center, LMU München, Munich, Germany). The following nChIP-seq protocol was adapted from (Punzeler et al., 2017) and (Link et al., 2018). Immunoprecipitations were carried out with S1 mononucleosomes derived from Pmi28 myoblasts and Pmi28 myotubes. nChIP-seq experiments were performed in independent duplicates.

### Sample preparation

#### Pmi28 myoblasts

Pmi28 myoblasts were washed with PBS (Sigma-Aldrich), detached using 0.25 % Trypsin/1 mM EDTA (Pan BioTech) for 3 min and collected in a centrifuge tube. After centrifugation for 4 min at 1300 g, cells were washed with PBS and centrifuged again. The pellet obtained from three 150 mm cell culture dishes (Corning) roughly corresponding to 15 Mio cells was snap frozen in liquid nitrogen.

#### Pmi28 myotubes

In order to separate differentiated myotubes from undifferentiated myoblasts, cells were harvested in the following way: first, myotubes were gently trypsinized using 0.025 % Trypsin/0.1 mM EDTA and progress of trypsinization was monitored with a phase contrast microscope; when myotubes detached, leaving the majority of myoblasts still attached to the bottom of the dish, the supernatant containing myotubes and some myoblasts was transferred to a fresh 150 mm cell culture dish (Corning). The remaining myoblasts re-adhered to the dish bottom within 30 to 60 min, whereas myotubes were not re-adherent within this time and could be collected with the supernatant yielding a pure myotube population. After centrifugation for 4 min at 1,300 g, cells were washed with PBS and centrifuged again. The pellet obtained from three 150 mm cell culture plates was snap frozen in liquid nitrogen.

### Micrococcal nuclease (MNase) digest of chromatin

Per immunoprecipitation, 5 Mio cells were used. Cells were resuspended in 750 µl PBS + 1 x PI + 0.3 % Triton X-100 and incubated for 10 min at 4 °C while rotating. The cell lysate was centrifuged for 10 min at 3,000 g and 4 °C, the pellet was resuspended in 625 µl PBS + PI and centrifuged for 5 min at 3,000 g at 4 °C. Supernatant was carefully removed and the pellet was resuspended in 62.5 µl EX100 (10 mM HEPES pH 7.6, 100 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 0.5 mM EGTA, 10 % (v/v) glycerol, 10 mM β-glycerol phosphate) + 1 x PI + 1 mM DTT and transferred into a DNA low binding tube (Eppendorf). CaCl<sub>2</sub> concentration was adjusted to 2 mM with 0.5 µl 250 mM CaCl<sub>2</sub> and 0.1875 µl MNase (1 U/µl, Sigma-Aldrich) were added. MNase digest was carried out for 20 min at 26 °C. The reaction was stopped by the addition of 1.25 µl 0.5 M EDTA to a final concentration of 10 mM and the suspension was centrifuged at 13,300 rpm for 30 min at

4 °C in a tabletop centrifuge. The supernatant (S1 mononucleosomes) was transferred into a fresh DNA low binding tube (Eppendorf) and filled up to 500 µl with EX100 + 1 x PI + 1 mM DTT. 25 µl (5 %) were saved as input.

### Chromatin Immunoprecipitation

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

### Washing

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

### DNA preparation

Washed beads were resuspended in 100 µl TE buffer. To the input DNA sample, 75 µl TE were added. 3 µl of 10 % SDS and 5 µl of proteinase K were added and the suspensions were incubated 1 h at 65 °C. After brief vortexing, suspensions were magnetically separated and supernatant was transferred into DNA low binding tubes (Eppendorf). Beads were washed with 100 µl TE + 0.5 M NaCl, vortexed briefly, magnetically separated and the supernatant added to the first supernatant. For input DNA, 100 µl TE + 0.5 M NaCl were added. Eluted IP DNA and input DNA were phenol/chloroform/isoamylalcohol extracted using MaXtract high density tubes (Qiagen) and ethanol precipitated according to a standard protocol. After extraction, the IP DNA pellets were resuspended in 12 µl 10 mM Tris-HCl pH 7.5 and the input DNA pellet was resuspended in 20 µl 10 mM Tris-HCl pH 7.5.

### Quality control

For quality control, DNA concentrations were determined via the Qubit dsDNA High Sensitivity Kit (Invitrogen) and DNA size was analyzed on a DNA 1000 BioAnalyzer chip (Agilent).

### Library preparation

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

### Sequencing

Sequencing was performed at the laboratory for functional genome analysis (LAFUGA, Gene Center, LMU München, Munich, Germany). ChIP-seq libraries were sequenced on an Illumina HiSeq 1500 platform using single read 50 nucleotide sequencing.

### Analysis

Read mapping and bioinformatic analyses were performed by Dr. Sebastian Bultmann (BioSysM, LMU München, Munich, Germany). ChIP-seq reads were aligned to the mouse genome (mm10) with Bowtie (v.1.2.2) with parameters '-a -m 3 -n 3 --best --strata'. Subsequent ChIP-seq analysis was carried out on data of merged replicates. Signal pile up was performed using MACS2 callpeak (Zhang et al., 2008) with the parameters '--extsize 150' for ChIP, '--extsize 220' for hMeDIP, and '--nomodel -B --nolambda' for all samples.

## 2.7 CHROMATIN CONFORMATION CAPTURE (HI-C)

### Sample preparation

Hi-C was performed as described previously (Belaghzal et al., 2017) by Dr. Yana Feodorova (Medical University of Plovdiv, Plovdiv, Bulgaria) with help of Dr. Erica Hildebrand (University of Massachusetts Medical School, Worcester, USA) in the laboratory of Prof. Dr. Job Dekker (University of Massachusetts Medical School, Worcester, USA). Three biological replicates were used for each of the studied cell types – myoblasts, myotubes, thyrocytes and bladder.

For each thyroid sample, thyroid glands from 6 mice were used. For each bladder sample, the bladder from one mouse was used. Glands and bladder were dissected, cleaned from neighboring tissues, minced with micro-scissors into small pieces and transferred to a 2 ml tube using low retention blue tips (BRAND). Fixation was carried out in 1 % formaldehyde for 10 min.

Sample preparation on myoblasts and myotubes for Hi-C was done in the same way as for ChIP-seq experiments, but cells were fixed (1 % formaldehyde for 10 min) and then washed once in PBS prior to snap freezing. Per sample, cells from three p150 dishes were used.

### Sequencing

Hi-C libraries were sequenced using Illumina 50 bp paired-end sequencing in the laboratory of Prof. Dr. Job Dekker (University of Massachusetts Medical School, Worcester, USA).

### Analysis

Hi-C data processing was conducted by Dr. Johannes Nübler (Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, USA) as described previously (Imakaev et al., 2012). Iterative correction was applied to the data in order to remove biases as much as possible (Imakaev et al., 2012). *Cis/trans* contact frequencies were calculated for each bin as the number of *cis* (same chromosome) contacts divided by the number of *trans* (different chromosome) contacts of that bin. For plotting, x-axes were rescaled for each gene, so that TSS and TTS align. Interpolation was used to yield the same number of points in each gene. Plotted quantities were normalized to regions outside the genes (the regions from  $[TSS - 3 * \text{gene length}]$  to  $[TSS - 0.5 * \text{gene length}]$  and  $[TTS + 0.5 * \text{gene length}]$  to  $[TTS + 3 * \text{gene length}]$  was used to normalize *cis/trans* profiles to unity). Long lowly expressed control genes for *cis/trans* ratio plots are listed in Appendix Table S8.

## 2.8 MICROSCOPY

### 2.8.1 LIGHT MICROSCOPY

#### Epifluorescent microscopy

To examine chromosomal location of BACs on 2D FISH preparations and to assess the quality of FISH and immunostaining, an epifluorescent microscope (Axiophot 2, Zeiss) was used. The system was equipped with a CCD camera (Coolview CCD Camera System) and the MetaVue software (Zeiss).

#### Confocal laser scanning microscopy

Image stack series 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 typically set to 300 nm. XY pixel size varied from 20 to 60 nm, depending on the final magnification. Used laser lines are depicted in Table 2.9.

**TABLE 2.9: Equipped laser lines on Leica SP5**

Fluorophore	Excitation	Emission	Laser line
<b>DAPI</b>	358 nm	461 nm	405 nm (Diode)
<b>FITC, Alexa488</b>	490 nm	525 nm	488 nm (Argon)
<b>Cy3</b>	554 nm	568 nm	561 nm (DPSS)
<b>TAMRA, Alexa555</b>	555 nm	580 nm	
<b>TexasRed</b>	596 nm	615 nm	594 nm (HeNe)
<b>Cy5</b>	649 nm	666 nm	633 nm (HeNe)

#### Processing of image stacks

Confocal stacks were processed using the ImageJ software (<https://imagej.nih.gov/ij/>). Before analysis, stacks were corrected for chromatic shift in z direction using the StackGroom/z-shift corrector plugin. This plugin re-aligns the images of an acquired z-stack according to the axial chromatic shift measured using TetraSpeck beads (Molecular Probes).

RGB image stacks with all desirable combinations of channels were generated using the StackGroom/3channels plugin.

Images in this study are maximum intensity projections of several shift corrected optical planes of a z-stack (typically covering 1 to 1.5  $\mu\text{m}$ ) unless stated otherwise.

## 2.8.2 ELECTRON MICROSCOPY

Blocks for sectioning were prepared in the laboratory of Prof. Dr. Andreas Klingl (Biocenter, LMU München, Munich, Germany). In brief, Pmi28 myoblasts and *in vitro* differentiated myotubes were grown on coverslips and fixed with 2.5 % glutaraldehyde in 300 mOsm cacodylate buffer (75 mM Cacodylate, 75 mM NaCl, 2 mM MgCl<sub>2</sub>), postfixed in 1 % osmium tetroxide and contrasted with 8 % uranyl acetate. Embedding was performed in Epon epoxy resin.

Sectioning of Epon-embedded cells, contrasting of sections with lead citrate after Reynolds and microscopy were performed in the laboratory of Dr. Yannick Schwab (European Molecular Biology Laboratory, Heidelberg, Germany), using a screening technique allowing high cell number acquisition.

## 2.9 BIOINFORMATICS

### 2.9.1 GTEx DATA ANALYSIS

GENCODE gene annotation version v19 (human) was downloaded from <https://www.genencodegenes.org/>. Using R (<https://www.R-project.org/>), mitochondrial genes were excluded from the analysis and only protein-coding genes with a known gene status were selected to ensure proper annotation and used for further evaluation. Gene lengths were calculated from GENCODE annotated start and end positions of each gene. RNA-seq data on human tissues was downloaded from the website of the GTEx portal (<https://gtexportal.org/home/datasets>, version 7, median gene level TPMs). GENCODE annotations were joined with GTEx RNA-seq data based on the genes' Ensembl gene ID.

### 2.9.2 SEMI-AUTOMATED GENE FLANKS MEASUREMENT

3D gene flanks distance measurements were conducted on confocal stacks using a semi-automated script developed by David Hörl (BioSysM, LMU München, Munich, Germany).

The signal spots in both channels were identified by detecting local minima in a Laplacian-of-Gaussian (LoG)-filtered image (with the expected size/sigma 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 step, results were visualized immediately, with the option to manually curate them and remove erroneous detections or 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](https://github.com/hoerldavid/fish_analysis).

### 2.9.3 COMPUTATIONAL CHROMATIN MODELING

Computational modeling was conducted by Dr. Johannes Nübler (Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, USA). Chromatin was modeled as a polymer with a monomer size of 1 kb roughly corresponding to 5 nucleosomes packed in a 20 nm globule and included 6 chromosomes, each 50 Mb in length. 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 mixed only moderately and retained their territoriality. For completeness, subdivision of chromatin into A- and B-compartments with attraction of B-compartments to the nuclear periphery were included. On each chromosome, a small region was assigned as our gene of interest, and it was explored how TL formation is reproduced by changing different parameters.

Simulations were based on polymer simulation code developed in the laboratory of Prof. Dr. Leonid Mirny (Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, USA). Polymer simulations were performed using a Mirny laboratory written wrapper (available at <https://github.com/mirnylab/openmm-polymer-legacy>) around the open source GPU-assisted molecular dynamics package OpenMM (Eastman et al., 2017). Polymers are represented as a chain of monomers with harmonic bonds, a repulsive excluded volume potential, and an additional small attraction for the interaction of two monomers of type B. To obtain Hi-C maps from simulated data, polymer conformations were first coarse grained by a factor of 10 (i.e. only every 10th monomer is considered) in order to reduce the size of the computed Hi-C matrix. Then, a cutoff radius was defined, mimicking the crosslinking radius in an actual Hi-C experiment. The cutoff radius was 10 monomer diameters (we verified that results are insensitive to the cutoff).



---

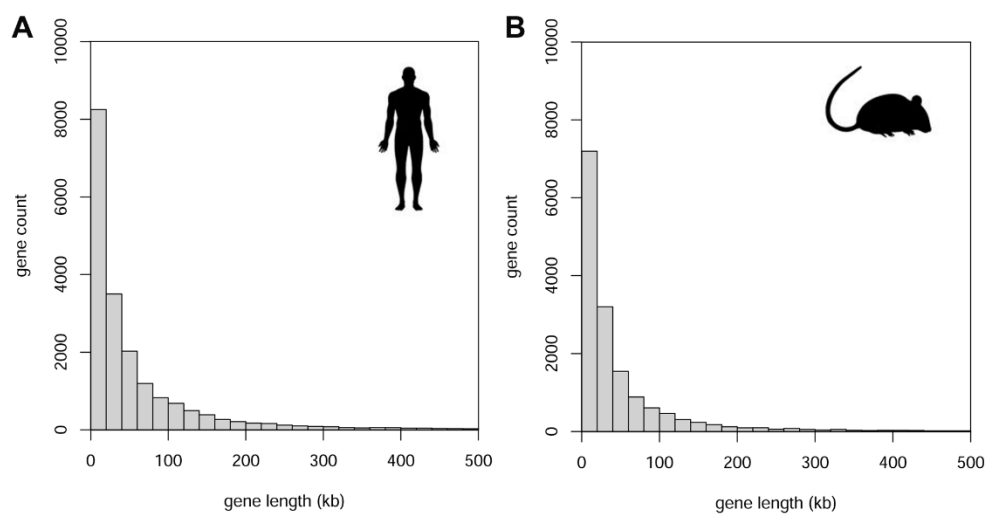
## 3 RESULTS

### 3.1 LONG MAMMALIAN GENES ARE RARELY HIGHLY EXPRESSED

Due to the resolution limit of light microscopy that prevents structural resolution of short lowly expressed genes, we first aimed at finding genes that are both long and highly expressed and thus allow their microscopic visualization.

#### 3.1.1 THE MAJORITY OF MAMMALIAN GENES IS SHORT

To estimate the amount of possible candidate genes for our subsequent studies, we analyzed the length of GENCODE annotated protein coding genes in human and mouse. In human, 43 % of the genes span below 20 kb and only 18 % of the genes span more than 100 kb (Figure 3.1 A). Mouse genes are on average smaller: 46 % of the genes measure below 20 kb and 14 % of the genes measure above 100 kb (Figure 3.1 B).



**FIGURE 3.1: The majority of human and mouse genes is short.** Histograms depicting the distribution of gene lengths in human and mouse. Genes were annotated according to GENCODE. **(A)** 43 % of human protein coding genes measure below 20 kb, 18 % measure more than 100 kb. **(B)** 46 % of mouse protein coding genes measure below 20 kb, 14 % measure more than 100 kb. *Only genes with a length below 500 kb are depicted. Bins represent 20 kb.*

### 3.1.2 LONG HUMAN GENES ARE RARELY HIGHLY EXPRESSED

In order to select candidate genes that are both long and highly expressed, we consulted a comprehensive published RNA-seq dataset from the Genotype Tissue Expression Portal (GTEx) for a variety of human tissues. In the dataset comprising 51 tissues, only ten protein coding genes exceeded both a gene length of 100 kb and a transcription level of 1000 transcripts per million (TPM) (Table 3.1) indicating that long genes are not only rare but also generally not highly expressed.

**TABLE 3.1: Ten genes > 100 kb and > 1000 TPM found in the GTEx dataset.**

Tissue	Gene(s)
Artery: Aorta	<i>MYH11, MYH9, RGS5, CALD1</i>
Artery: Coronary	<i>MYH11</i>
Artery: Tibial	<i>MYH11, RGS5, CALD1, MYH9</i>
Bladder	<i>MYH11</i>
Brain: Hippocampus	<i>MBP</i>
Brain: Spinal cord cervical	<i>MBP</i>
Brain: Substantia nigra	<i>MBP</i>
Ectocervix	<i>MYH11</i>
Colon sigmoid	<i>MYH11, CALD1</i>
Colon transverse	<i>MYH11</i>
Esophagus: Gastroesophageal junction	<i>MYH11</i>
Esophagus: Mucosa	<i>SPINK5, JUP</i>
Esophagus: Muscularis	<i>MYH11, CALD1</i>
Fallopian Tube	<i>MYH11</i>
Skeletal Muscle	<i>MYBPC1</i>
Prostate	<i>MYH11</i>
Skin: not sun exposed	<i>JUP</i>
Skin: sun exposed	<i>JUP</i>
Thyroid	<i>TG, TPO</i>
Uterus	<i>MYH11</i>

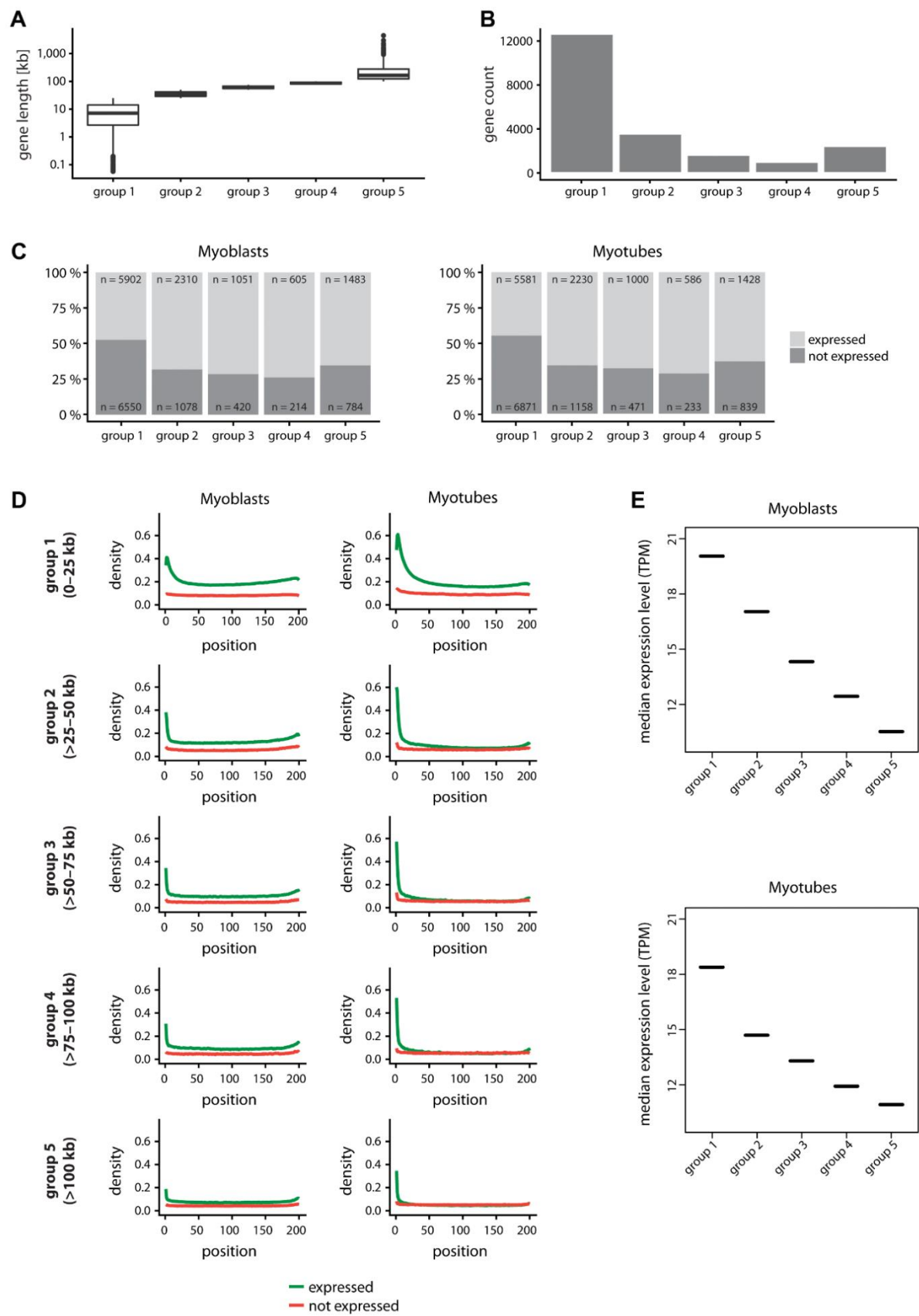
### 3.1.3 LONG GENES EXHIBIT LOWER POLYMERASE LEVELS THAN SHORT GENES

Another line of evidence for the low expression of long genes arose from our own ChIP-seq experiments. We conducted native ChIP-seq with an antibody against the CTD of RNA Pol II in undifferentiated myoblasts and differentiated myotubes in order to compare RNA Pol II occupancy of genes depending on their length and expression level. For evaluation, genes were divided into five size groups (group 1: 0–25 kb; group 2: > 25–50 kb; group 3: > 50–75 kb; group 4: > 75–100 kb; group 5: > 100 kb) (Figure 3.2 A). In accordance with the fact that the majority of mammalian genes is short, we found group 1 to contain the most genes (Figure 3.2 B). We further divided genes into “expressed” (> 1 TPM) and “not expressed” (< 1 TPM) in the respective cell type (Figure 3.2 C). The number of genes in each group was highly similar between myoblasts and myotubes, indicating that a similar number of genes of the respective length is expressed in both cell types. However, average polymerase occupancy along gene bodies varied in the different groups and was directly proportional to the gene length: the shorter the genes, the higher the polymerase occupancy (Figure 3.2 D). This was also supported by the finding that smaller genes on average had a higher transcription level than long genes indicated by higher average TPM values as determined by RNA-seq (Figure 3.2 E). Collectively, these data provide further evidence that long genes are on average lower expressed than short genes.

(Figure on the next page)

**FIGURE 3.2: Expressed short genes exhibit a higher RNA Pol II occupancy than expressed long genes.** ChIP-seq with an antibody against the CTD of RNA Pol II in Pmi28 mouse myoblasts and differentiated myotubes showed that expressed genes overall display a higher occupancy with RNA Pol II along their gene bodies than non-expressed genes in both cell types. The difference is especially prominent at the TSS. In the group of expressed genes, the RNA Pol II occupancy of gene bodies is on average higher in small genes than in long genes. **(A)** Genes were divided into five groups depending on their size (group 1: 0–25 kb, group 2: > 25–50 kb, group 3: > 50–75 kb, group 4: > 75–100 kb, group 5: > 100 kb). **(B)** Group 1 is by far the biggest group (12,452 genes), followed by groups 5 (2,267 genes), 2 (3,388 genes), 3 (1,471 genes), and 4 (819 genes). **(C)** Genes in the different size groups were divided according to their transcriptional status into expressed (>1 TPM, light gray) and non expressed (<1 TPM, dark gray) genes. **(D)** RNA Pol II density is generally higher around the TSS compared to the gene body in both cell types (left: myoblasts; right: myotubes) and irrespective of the gene length. Pol II density along the gene bodies of transcribed genes (green) correlated with gene length: the smaller the genes, the higher the Pol II density. For non-expressed genes (red), Pol II levels were generally low and did not correlate with gene length. *Each gene was divided into 200 equally sized bins (x-axis, position) and genes were aligned according to bins.* **(E)** The median expression levels of genes belonging to the “expressed” group were higher in groups containing smaller genes, and negatively correlated with gene length in myoblasts (top) as well as in myotubes (bottom): The longer the genes, the lower the median expression level. Expression levels (median): Myoblasts: group 1: 20.1 TPM; group 2: 17 TPM; group 3: 14.3 TPM; group 4: 12.4 TPM; group 5: 10.5 TPM. Myotubes: group 1: 18.3 TPM; group 2: 14.7 TPM; group 3: 13.3 TPM; group 4: 11.9 TPM; group 5: 10.9 TPM. *Data from RNA-seq experiments, see Results section 3.1.5.*

RESULTS



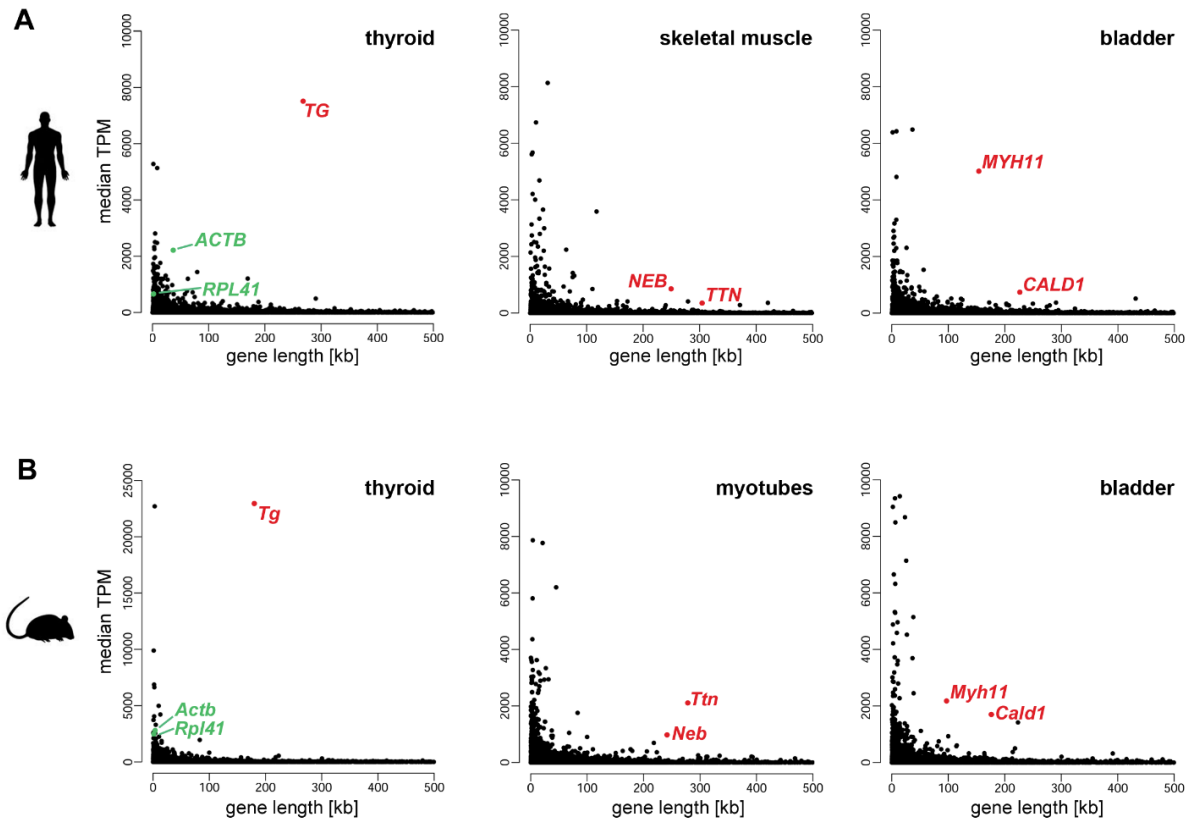
### 3.1.4 SELECTION OF LONG HIGHLY EXPRESSED HUMAN GENES FROM GTEx DATASET

For our studies, we aimed at investigating long and highly expressed genes coding for structural proteins and expressed in cell types that are easily identifiable on tissue sections to enable their visualization in mouse tissues. To find such genes, we mined the GTEx RNA-seq dataset for possible candidates. One of the human genes with a high expression level is thyroglobulin (*TG*) (Table 3.1), expressed in the thyroid gland and coding for a protein serving as a precursor for thyroid hormones triiodothyronine (T3) and thyroxine (T4) (Di Jeso and Arvan, 2016). In humans, the *TG* gene measures  $\approx 268$  kb and is expressed at 7,510 TPM, strongly exceeding the expression levels of all other protein-coding genes in the thyroid gland, including housekeeping genes, such as beta actin (*ACTB*; 2,214 TPM) and ribosomal protein L41 (*RPL41*; 652 TPM) by 3.4- and 11.5-fold, respectively (Figure 3.3 A).

In addition to *TG*, we selected four genes specifically expressed in muscle cells: the smooth muscle isoform of myosin heavy chain 11 (*MYH11*,  $\approx 154$  kb and 5,006 TPM) and caldesmon 1 (*CALD1*,  $\approx 226$  kb and 736 TPM) highly expressed in smooth muscle, as well as titin (*TTN*,  $\approx 305$  kb and 350 TPM) and nebulin (*NEB*,  $\approx 249$  kb and 858 TPM) highly expressed in skeletal muscle (Figure 3.3 A). *MYH11* codes for a smooth muscle specific contractile protein belonging to the myosin heavy chain family and is capable of hydrolyzing adenosine triphosphate (ATP) and thereby converting chemical energy to mechanical energy (Weiss and Leinwand, 1996). *CALD1* encodes a calmodulin- and actin-binding protein and inhibits the ATPase activity of myosin (Huber, 1997). *TTN* encodes the largest human protein that is essential for the stability, elasticity and passive tension of striated muscle (Linke, 2018). *NEB* codes for an actin-binding filamentous protein and is part of the skeletal muscle thin filament (Chu et al., 2016).

### 3.1.5 VALIDATION OF SIMILAR CHARACTERISTICS FOR ORTHOLOGOUS MOUSE GENES

To confirm the high expression of the five selected genes in the corresponding mouse tissues, we performed RNA-seq of thyroid and smooth muscle tissue, as well as cultured myoblasts and myotubes differentiated from myoblasts *in vitro*. Expression of the *Tg* gene ( $\approx 180$  kb) in the mouse thyroid was exceptionally high (22,924 TPM), exceeding the expression levels of highly and robustly expressed housekeeping genes by almost 10 fold (e.g. *Actb*  $\approx 4$  kb and 2,791 TPM; *Rpl41*  $\approx 1$  kb and 2,467 TPM) (Figure 3.3 B left). RNA-seq analysis of bladder smooth muscle tissue confirmed the high expression level of *Myh11* ( $\approx 97$  kb and 2,180 TPM) and *Cald1* ( $\approx 177$  kb and 1,698 TPM) in the mouse (Figure 3.3 B right). The expression levels of the genes of interest in myotubes were compared to the expression levels in undifferentiated myoblasts. Both genes (*Ttn*  $\approx 279$  kb and 2,100 TPM and *Neb*  $\approx 242$  kb and 980 TPM) exhibited a high expression in myotubes (Figure 3.3 B middle) and low expression in myoblasts (both genes  $\approx 30$  TPM). Of note, the expression of the genes in myoblasts above the zero level can be explained by a heterogeneous population of myoblasts, some of which start spontaneous differentiation into myotubes.



**FIGURE 3.3: Five genes are long and highly expressed in both human and mouse tissues.** Scatterplots showing the gene length (x-axis) and the transcription level (in TPM, y-axis) of all genes below 500 kb detected by RNA-seq in the respective tissue in (A) human and (B) mouse. Selected genes (red) stick out in terms of gene length and transcription level in both species and show substantially higher expression rates than exemplary housekeeping genes (green). Human RNA-seq data taken from GTEx.

## 3.2 EXPRESSED LONG GENES FORM TRANSCRIPTION LOOPS

Next, we sought to visualize the five candidate genes detected by RNA-seq via Fluorescence in situ Hybridization (FISH) on tissue cryosections or on cultured cells. We reasoned that – due to their high expression level – these genes should be densely decorated with nascent RNA transcripts. Similar to lateral loops of lampbrush chromosomes, the combination of a high amount of nascent RNA transcripts and the large size of the selected genes should enable their visualization with light microscopy.

### 3.2.1 FISH PROBE DESIGN

FISH probes, if not indicated otherwise, were BAC clones labeled by NT with either fluorophores or haptens (see Materials and Methods section 2.4.3). BAC derived probes span large regions ( $\approx 150$ – $250$  kb) and, importantly, they are hybridizing mostly to introns because they are generally longer than exons (Deutsch and Long, 1999; Sakharkar et al., 2004; Zhu et al., 2009). Therefore, for visualizing DNA or RNA of a long gene, BAC probes are well suited. However, for labeling of short regions and specifically for labeling of selected introns or exons, a proportion of probes was generated by PCR from cDNA and genomic templates, or oligo probes were used (see Materials and Methods section 2.4.4).

### 3.2.2 3D FISH ON LONG HIGHLY EXPRESSED GENES REVEALS TRANSCRIPTION LOOPS

To visualize the genes selected from RNA-seq data, we carried out both DNA and RNA FISH on cryosections of the corresponding mouse tissues or cultured cells using BAC probes (Figure 3.4 A). Additionally, we visualized two “control genes” – highly expressed but short actin alpha 1 (*Acta1*) and long but lowly expressed dystrophin (*Dmd*).

In DNA FISH experiments, we observed two different FISH signal patterns corresponding to the gene’s expression status. Expressed genes exhibited a dispersed FISH signal, sometimes consisting of several small foci in the nuclear interior (Figure 3.4 C). In contrast, silent genes assumed a dot-like shape with high fluorescence intensity and resided at the nuclear periphery or chromocenter surface (Figure 3.4 B). These different patterns suggest that transcribed long genes are highly decondensed, whereas they are condensed in a silent state.

RNA FISH with the same set of probes did not show signals in cells in which the respective gene is silent. Expressing cells, in contrast, yielded RNA FISH signals in the shape of coiled loops in all tested genes (Figure 3.4 D). These loops were especially large (up to  $10\ \mu\text{m}$ ) in case of highly upregulated *Tg*. Importantly, nuclei exhibited two loops corresponding to two alleles (Figure 3.4 D).

Due to the large length of introns (Deutsch and Long, 1999; Sakharkar et al., 2004; Zhu et al., 2009), the BAC derived genomic FISH probes mainly label intronic regions in nascent RNA transcripts (Figure 3.4 A). RNA FISH signal size therefore mostly depends on the presence of and corresponds to the amount of nascent transcripts. Hence, we concluded that the loops formed by transcribed genes are microscopically visible because the gene bodies are covered with thousands of nascent transcripts, comparable to transcription units

## RESULTS

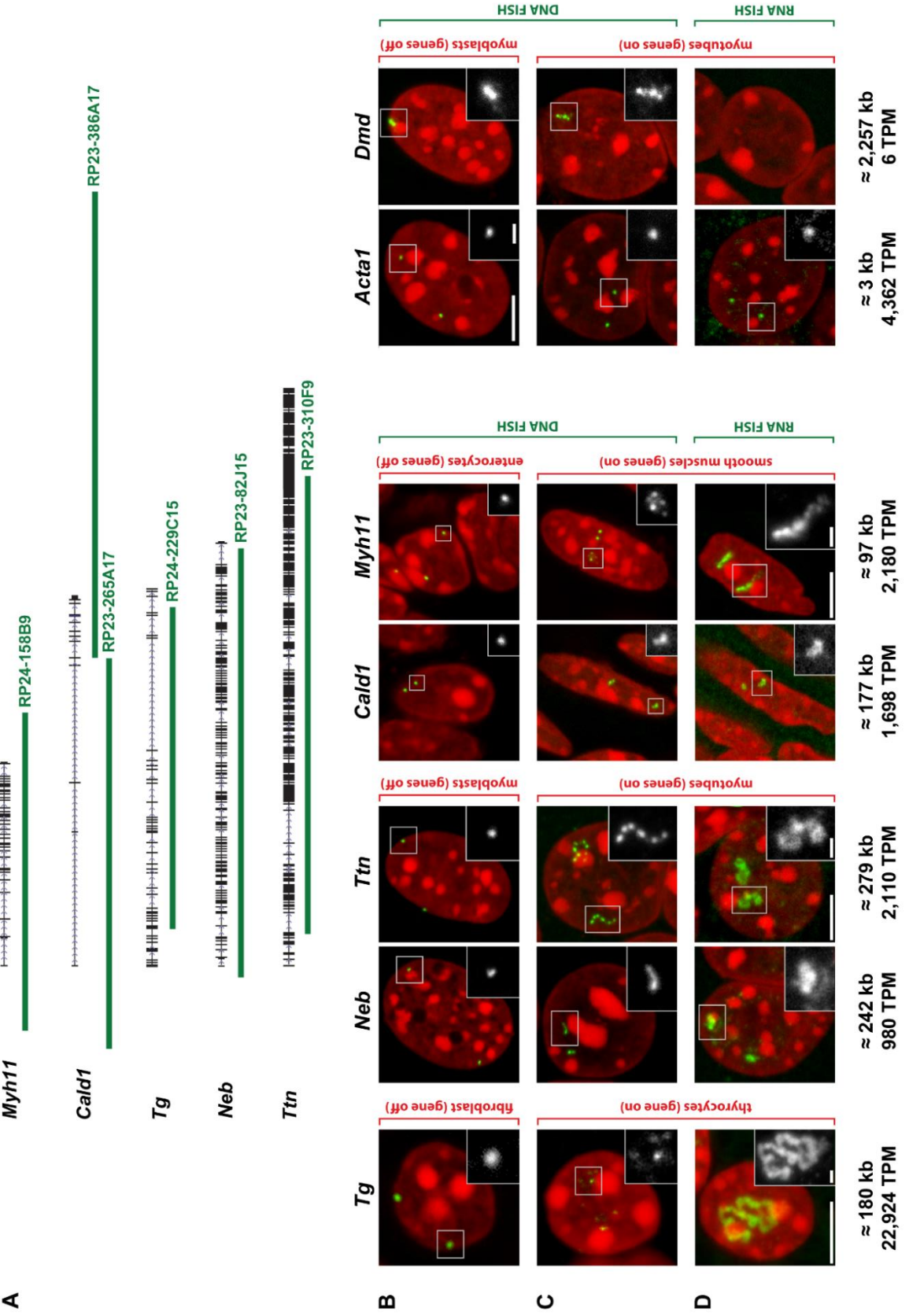
---

of lampbrush chromosomes and genes in polytene chromosome puffs. Thus, we termed these loops “Transcription Loops” (TLs). The data further corroborated our hypothesis that in order to be resolved by light microscopy, a gene requires both a reasonable length and a high transcription level. Consistent with that, the highly expressed (4,362 TPM) but short ( $\approx 3$  kb) *Acta1* gene did not form a TL in myotubes (Figure 3.4 D). Likewise, the exceptionally long but very lowly expressed (5.5 TPM) *Dmd* gene ( $\approx 2,257$  kb) also did not form a TL (Figure 3.4 D).

---

**FIGURE 3.4: Microscopic visualization of transcription loops formed by long and highly expressed genes.** (A) Green bars represent locations of genomic BAC probes relative to respective genes. Exons are depicted as vertical lines. BAC clone identifiers are indicated next to the bars. (B) DNA FISH in cells not expressing the respective gene, showed dot-like signals at the nuclear periphery or next to chromocenters. (C) DNA FISH in cells expressing the respective gene, revealed strongly decondensed gene bodies, especially prominent in the case of *Ttn*. For the short *Acta1* gene, gene body decondensation was not resolvable. (D) RNA FISH in cells expressing the respective genes displayed coiled loop-like structures of different sizes for *Tg*, *Neb*, *Ttn*, *Cald1* and *Myh11*. As in DNA FISH, the structure of transcribed *Acta1* was irresolvable. The very lowly expressed *Dmd* gene did not show an RNA FISH signal. DNA was counterstained with DAPI (red). Scale bars 5  $\mu\text{m}$  and 1  $\mu\text{m}$  (zoom).

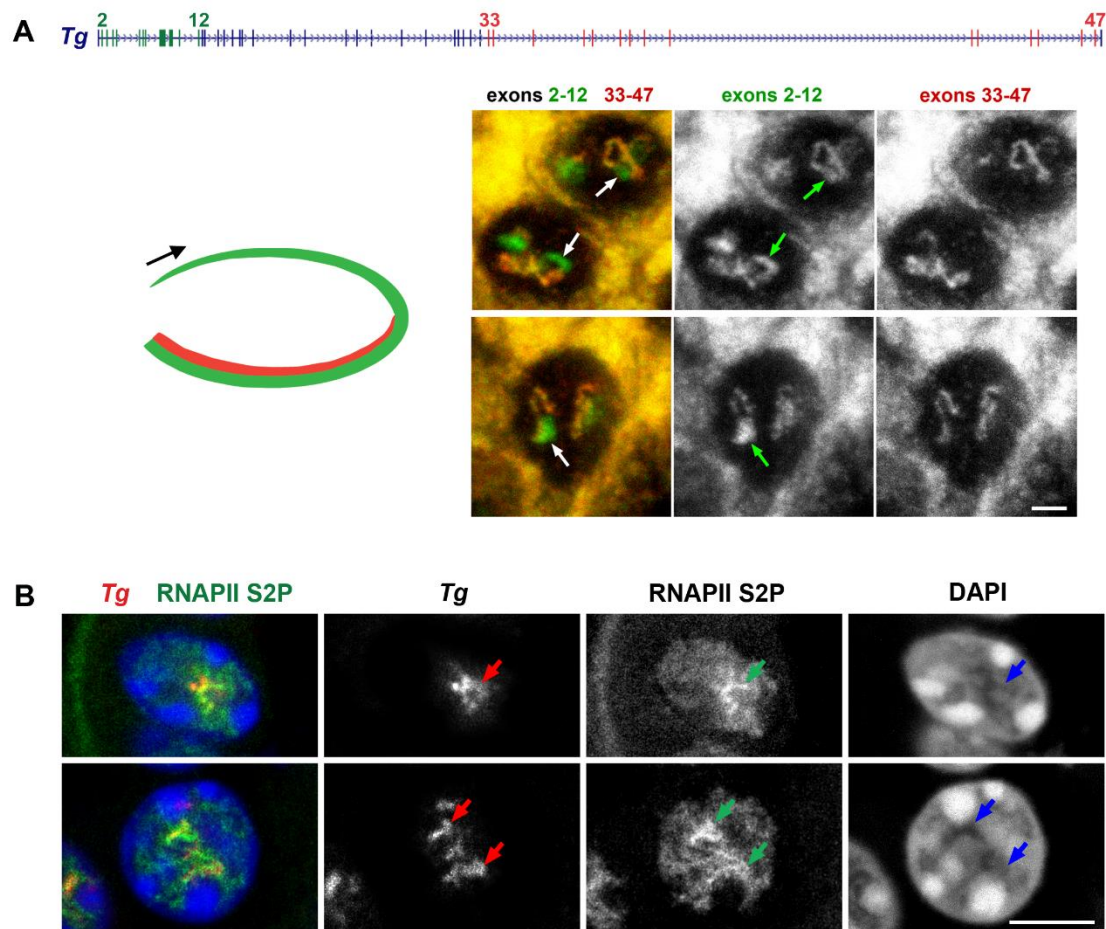




### 3.2.3 PROGRESSION OF TRANSCRIPTION ALONG TLs: VISUALIZATION OF EXONS AND ELONGATING POLYMERASES

To investigate whether the observed RNA FISH signals (Figure 3.4 D) represent nascent RNAs produced by ongoing transcription, we performed FISH experiments with probes differentially labeling specific exons at the 5' and 3' ends of the *Tg* gene. The two probes showed different labeling patterns: the 5' exon probe (exons 2–12) labeled the entire loop, whereas the 3' exon probe (exons 33–47) labeled only a part of the loop (Figure 3.5 A).

In contrast to introns, exons remain in the nascent RNA from the time they are transcribed. Therefore, early 5' exons mark transcripts along the whole TL, whereas 3' exons that are transcribed later mark transcripts deriving from later parts of the gene. Given the design of FISH probes, the sequential labeling of the TLs thus depends on the presence of exons in nascent RNAs and therefore reflects ongoing transcription. In line with that, we showed co-localization of the *Tg* TLs with accumulations of the elongating form of RNA Pol II (RNAPII S2P) (Figure 3.5 B). Folded TLs formed structures with such a high concentration of RNA and proteins that nucleoplasmic areas corresponding to the loops were visibly depleted of chromatin, as indicated by DAPI staining (Figure 3.5 B).

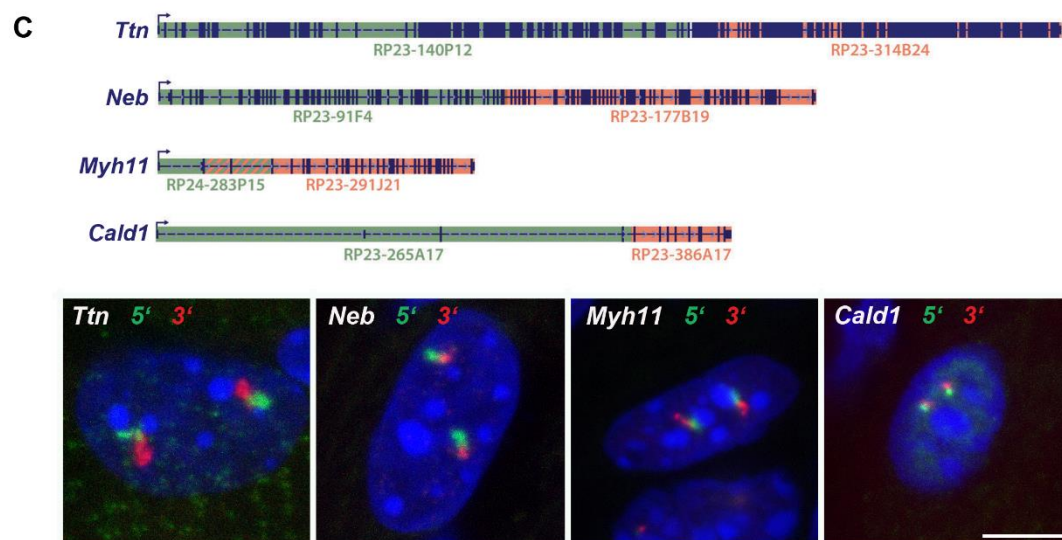
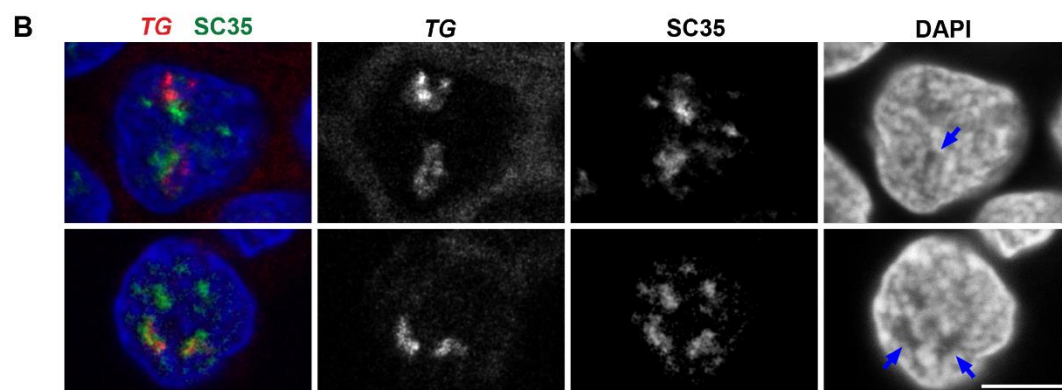
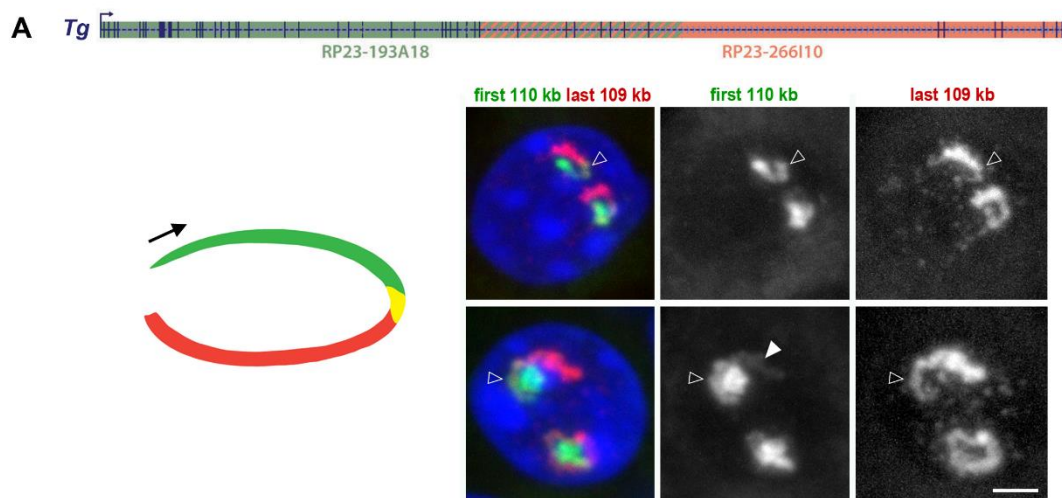


**FIGURE 3.5: Transcription loops manifest transcription progression. (A)** RNA FISH with probes specifically labeling exons at the 5' (exons 2–12, green) and 3' (exons 33–47, red) ends of *Tg* allows to follow transcription progression along transcription loops. Early transcribed exons (5') label the whole loop structure, whereas late transcribed exons (3') only label part of the loop. Top: Scheme of *Tg* gene highlighting the positions of exon probes generated by PCR. Exons are depicted as vertical lines. Bottom left: Schematic representation of the observed labeling pattern. Arrow indicates the direction of transcription. Bottom right: Examples of thyrocyte nuclei fully labeled by a probe for exons 2–12 (green) and partly labeled by a probe for exons 33–47. The early portion of the gene only labeled by the probe for exons 2–12 is indicated by white and green arrows. Scale bar 2  $\mu\text{m}$ . **(B)** *Tg* transcription loops (red and red arrows) co-localized with actively elongating RNA Pol II phosphorylated at CTD serine 2 (RNAPII S2P, green and green arrows). Co-localization occurred in euchromatic regions as indicated by DAPI staining (blue arrows). Scale bar 5  $\mu\text{m}$ .

### 3.2.4 PROGRESSION OF TRANSCRIPTION ALONG TLs: VISUALIZATION OF INTRONS AND NUCLEAR SPECKLES

In contrast to exon specific probes, RNA FISH with BAC probes labels mostly intronic regions, largely exceeding exons by length (Deutsch and Long, 1999; Sakharkar et al., 2004; Zhu et al., 2009) and therefore comprising most of nascent transcripts. Therefore, we used two sequential overlapping BACs to highlight *Tg* introns. The two probes labeled the *Tg* TL sequentially with a small portion of signal overlap (Figure 3.6 A). The 5' probe strongly labeled the first half of the loop decorated by nascent RNAs containing both unspliced introns and exons and only faintly labeled the second half of the loop with remaining 5' exons but spliced out 5' introns. The 3' probe marked the second half of the loop by hybridizing to both 3' exons and 3' introns. This successive pattern of labeling demonstrates co-transcriptional splicing occurring on *Tg* TLs. Furthermore, we confirmed the occurrence of co-transcriptional splicing for all of the selected long genes using differentially labeled genomic probes for 5' and 3' halves of the genes (Figure 3.6 C). Consistent with ongoing high splicing activity on the abundant transcripts, the *Tg* TLs were found either in close proximity to or overlapping with nuclear speckles (Figure 3.6 B). Notably, RNA FISH with genomic probes also labeled multiple granules in the nucleoplasm (Figure 3.6 A and 3.6 C), which we interpreted either as accumulations of spliced out but not degraded introns or as not yet exported mRNAs.

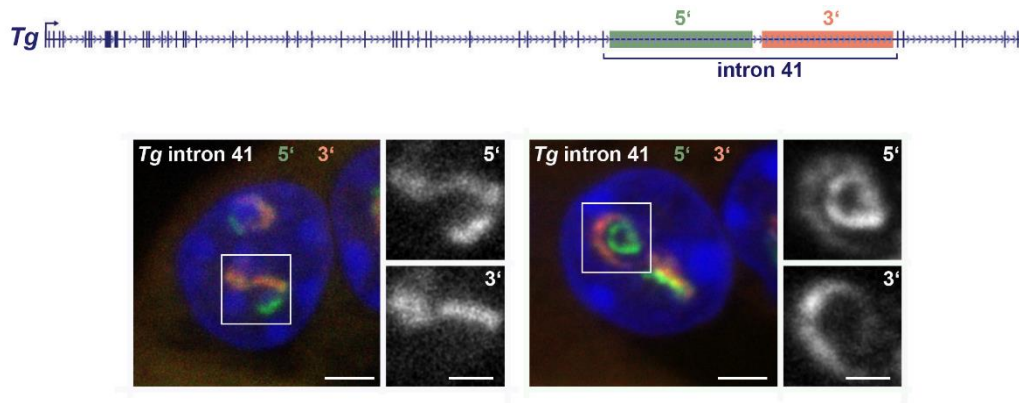
**FIGURE 3.6: Transcription loops exhibit co-transcriptional splicing.** (A) Genomic BAC probes (red and green) sequentially labeled the transcription loop due to hybridization to intronic sequences in nascent RNA (RNA FISH). Top: Scheme of *Tg* gene highlighting the positions of genomic BAC derived probes. *Probe overlap is indicated by diagonal stripes. Exons are depicted as vertical lines.* Bottom left: Schematic representation of the observed labeling pattern. *Arrow indicates the direction of transcription.* Bottom right: Examples of thyrocyte nuclei exhibiting sequential labeling pattern of the *Tg* transcription loop with BAC probes. The first half of the loop was labeled by the 5' probe (green), the second half of the loop was labeled by the 3' probe (red) due to hybridization to nascent RNA transcripts. The 5' probe yielded a faint signal in the second half of the loop due to hybridization to early exons (filled arrowheads). As BAC probes are overlapping, the mid part of the gene yielded signal from both probes (empty arrowheads). *Scale bar 2  $\mu$ m.* (B) *Tg* transcription loops (red) co-localized with (top) or resided in close proximity to (bottom) SC35 (green), a major constituent of nuclear speckles, in human thyrocytes. Transcription loops were typically localized in euchromatic regions (blue arrows). *Scale bar 5  $\mu$ m.* (C) Top: Schemes of *Ttn*, *Neb*, *Myh11* and *Cald1* genes highlighting the positions of genomic BAC derived probes. *Exons are depicted as vertical lines.* Bottom: Sequential labeling of transcription loops with two BAC probes was also observed for *Ttn* and *Neb* in mouse skeletal muscle and for *Myh11* and *Cald1* in bladder smooth muscle tissue. *DNA was counterstained with DAPI (blue).* *Scale bar 5  $\mu$ m.*





### 3.2.5 ABSENCE OF RECURSIVE SPLICING FOR THE LONGEST *Tg* INTRON

It has been observed that a number of introns is not excised as a whole but undergoes so-called recursive splicing – successive rounds of splicing of parts of the intron until the whole intron is removed (Georgomanolis et al., 2016; Sibley et al., 2015). Since the *Tg* gene harbors a large intron (intron 41) of 53.8 kb, we set out to investigate whether this intron is subject to recursive splicing. To this end, we generated two FISH probes for the two halves of intron 41, each spanning  $\approx 25$  kb. The probe hybridizing to the 5' half of the intron labeled the TL region corresponding to the entire intron, the probe for the 3' half of the intron labeled only half of this region (Figure 3.7). In case of recursive splicing, we would have expected the 5' half probe to label only a part of the intron as a result of co-transcriptional excision before the entire intron is transcribed. Absence of sequential labeling of the intron suggests that the 53.8 kb intron 41 of the *Tg* gene is not spliced recursively.



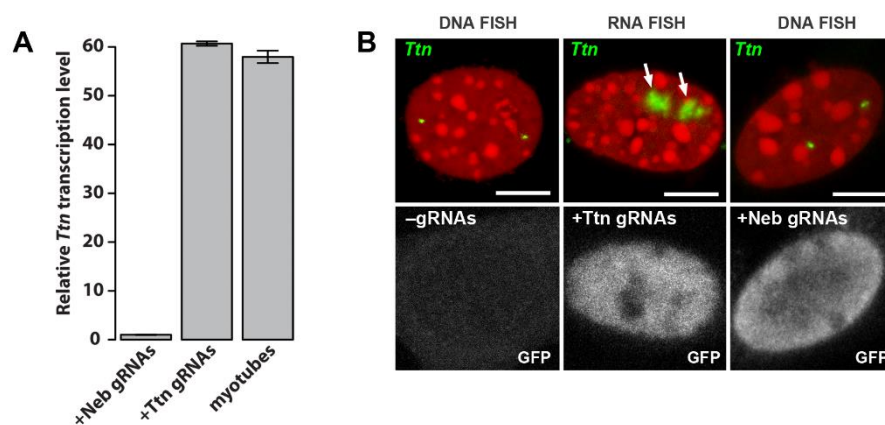
**FIGURE 3.7: The 53.8 kb *Tg* intron 41 is not spliced recursively.** Top: Scheme of *Tg* gene highlighting the location of PCR derived probes in intron 41. Each probe covers  $\approx 25$  kb. Exons are depicted as vertical lines. Bottom: Two examples of thyrocyte nuclei exhibiting labeling of the entire intron 41 by a probe for the first intron half (5', green) and labeling of only a part of intron 41 with a probe for the second half (3', red). DNA was counterstained with DAPI (blue). Scale bars 2  $\mu\text{m}$  and 1  $\mu\text{m}$  (zoom).

### 3.3 INDUCTION AND INHIBITION OF TRANSCRIPTION LOOPS

To provide further evidence that TLs are dynamic structures formed in response to transcriptional activation, we performed several experiments to either activate or inhibit transcription of the *Ttn* gene *in vitro*.

#### 3.3.1 INITIATION OF *Ttn* TRANSCRIPTION LOOPS BY TRANSCRIPTION INITIATION

First, we aimed at ectopically inducing *Ttn* TL formation in cells that do not express the gene endogenously. To this end, we generated myoblasts stably expressing a nuclease dead Cas9 (dCas9) conjugated with the transcription activator VP64-p65-Rta (VPR) (Chavez et al., 2015). VPR consists of three parts: 1) VP64, four tandem copies of VP16, typically inducing immediate early gene transcription in herpes simplex viruses, 2) p65, a TF involved in immune modulation, and 3) Rta, a TF of the Epstein Barr virus expressed during the immediate early stage of the lytic cycle. These modified myoblasts were co-transfected with a cocktail of six plasmids expressing gRNAs targeting the *Ttn* (Ttn gRNAs) or *Neb* (Neb gRNAs) promoter region and H2A-GFP that enables selection of transfected cells for qPCR and microscopic analyses. Upon transfection, firstly, *Ttn* transcription was induced to a level comparable to myotubes, exceeding the expression of *Ttn* in myoblasts transfected with Neb control gRNAs by > 60-fold (Figure 3.8 A). Secondly, 90 % of cells (n = 49) transfected with gRNAs targeting the *Ttn* promoter region exhibited *Ttn* TLs (Figure 3.8 B), indicating that TLs are formed upon transcription initiation.



**FIGURE 3.8: Transcription induction leads to the formation of *Ttn* TLs in mouse myoblasts.** Upon transcription induction via dCas9-VPR, transcription of the *Ttn* gene is upregulated and formation of a transcription loop is induced in Pmi28 mouse myoblasts. **(A)** The relative transcription levels of *Ttn* in Pmi28 myoblasts transfected with Neb control gRNAs or Ttn gRNAs and in differentiated myotubes were assessed via qPCR. *Ttn* transcription was induced in cells transfected with Ttn gRNAs to a comparable level as in myotubes corresponding to a > 60 fold induction compared to myoblasts transfected with Neb control gRNAs. Bars represent standard deviation. **(B)** Whereas untransfected control cells (-gRNAs) as well as cells transfected with control off-target gRNAs (+Neb gRNAs) did not show transcription loop formation of the *Ttn* gene, but revealed condensed gene bodies (top), cells transfected with Ttn gRNAs (+Ttn gRNAs) exhibited TL formation in both alleles in 90 % of cells (arrows). DNA was counterstained with DAPI (red). Scale bar 5  $\mu$ m.

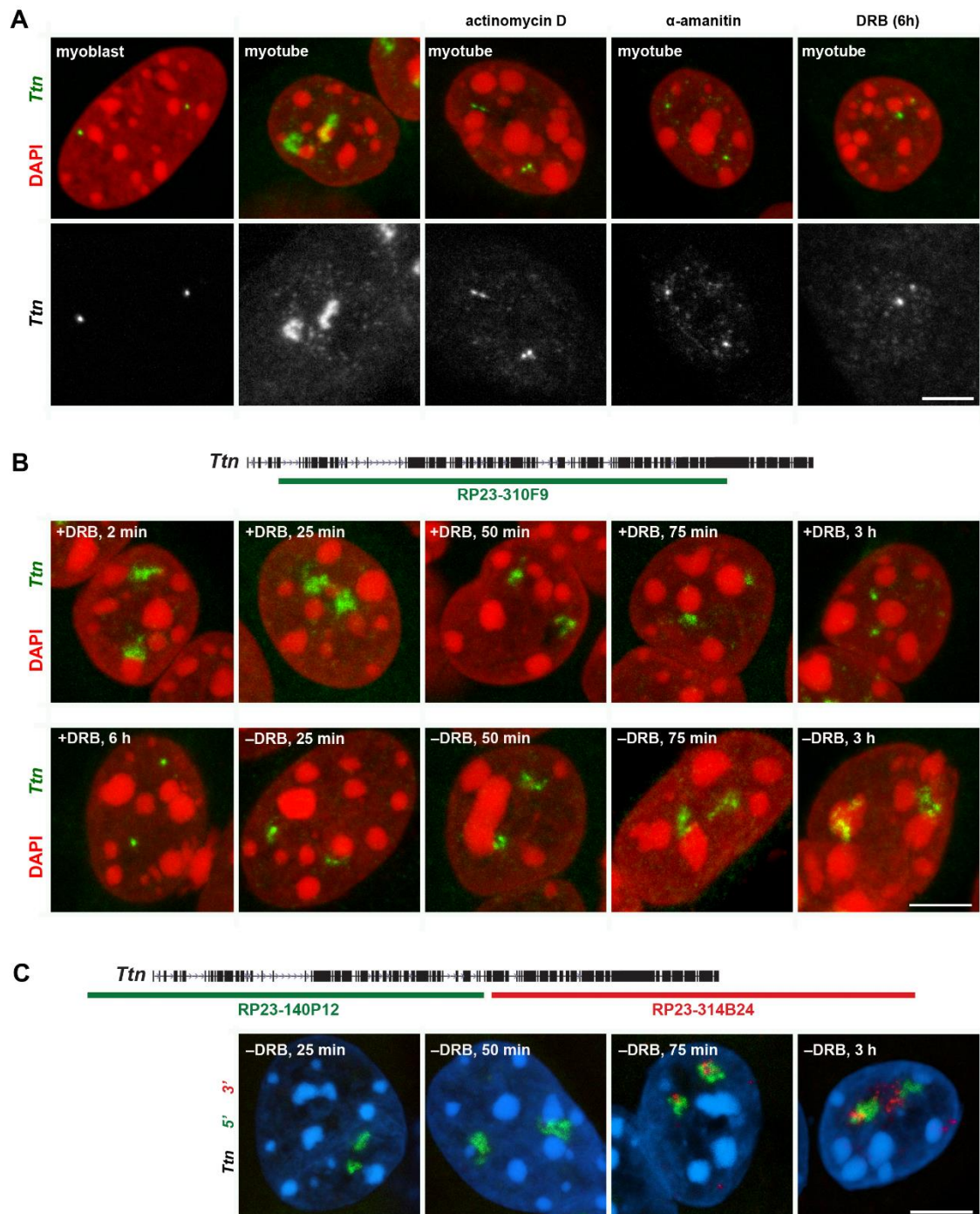
### 3.3.2 ELIMINATION OF *Ttn* TRANSCRIPTION LOOPS BY TRANSCRIPTION INHIBITION

Second, we aimed at eliminating *Ttn* TLs by transcription inhibition. We applied three different inhibitors with various mechanisms. 1)  $\alpha$ -Amanitin is a cyclic peptide derived from the mushroom genus *Amanita* and selectively inhibits RNA Pol II and III. It interferes with the bridge helix between the two largest Pol II subunits, Rpb1 and Rpb2, and thus hinders DNA and RNA translocation, slowing down transcription to a few bp per minute (Bushnell et al., 2002; Rudd and Luse, 1996). 2) Actinomycin D is a cyclic peptide antibiotic that binds directly to DNA by intercalating into GC rich regions, thereby stabilizing covalent complexes with topoisomerase I and preventing RNA polymerase progression at the initiation site (Sobell, 1985; Trask and Muller, 1988). 3) The adenosine analog 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) reversibly inhibits CDK9 and therefore prevents the phosphorylation of negative elongation factor (NELF) and DRB sensitivity inducing factor (DSIF), hampering Pol II from escaping the promoter into active elongation and holding it in a stalled state a few bp downstream of the TSS (Nechaev and Adelman, 2011).

Transcription inhibition with either of these drugs abolished the formation of TLs (Figure 3.9 A). A dot-like bright FISH signal indicated that the gene body condenses upon drug treatment (Figure 3.9 A). Applying DRB, reversibly keeping Pol II in a stalled state, allowed to monitor the kinetics of TL formation. Upon DRB treatment, no polymerases are released into elongation, whereas the already elongating polymerases continue to move along the gene body. Consistent with that, upon incubation with DRB, the RNA FISH signal was gradually reducing, reaching its minimum size after 3 hours of incubation (longer incubation for up to 6 hours did not further diminish the signal). After drug withdrawal, the RNA FISH signal was restored and gradually grew over time (Figure 3.9 B). Within 75 min after drug withdrawal, *Ttn* TLs reached their full size in some myotube nuclei. Given an elongation speed of RNA Pol II of 3.8 kb/min (Singh and Padgett, 2009), this is consistent with the estimated time for one full round of transcription of the *Ttn* gene of approx. 75 min.

In addition, using differentially labeled probes for 5' and 3' nascent RNAs, we demonstrated that after DRB removal the 5' RNA signal (corresponding to early RNA transcripts) emerges and grows first, whereas the 3' signal (corresponding to late RNA transcripts) emerges and grows later (Figure 3.9 C). These data provide further evidence that TL formation depends on active transcription.





**FIGURE 3.9: Transcription loop formation is obliterated by transcription inhibition.** (A) Transcription loop formation of *Ttn* is abolished upon treatment with either of the transcriptional inhibitors actinomycin D,  $\alpha$ -amanitin or DRB. FISH labeling both DNA and RNA on drug treated Pmi28 myotubes revealed dot-like signals corresponding to a condensed gene body, comparable to the signal observed in myoblasts (left). Scale bar 5  $\mu$ m. (B) Treatment of myotubes with DRB led to a gradual shrinkage of the *Ttn* TL as visualized with a genomic BAC probe (scheme, green bar) by FISH labeling both DNA and RNA. Upper panel: *Ttn* TLs were withdrawn gradually (left to right) as drug treatment time increases (+DRB). After 3 h of treatment, the gene body appeared as condensed bright spot. Lower panel: Starting from a condensed gene body after 6 h of DRB treatment, removal of the drug (-DRB) led to a gradual restoration of the *Ttn* transcription loop (left to right). BAC clone identifier is indicated below green bar. DNA was counterstained with DAPI (red). Scale bar 5  $\mu$ m. (C) RNA FISH using differentially labeled probes for the 5' and 3' regions of *Ttn* (scheme, green and red bars) revealed that after drug removal, the 5' signal (green) grew first and reached its full DRB). Scale bar 5  $\mu$ m.

### 3.4 CORRELATION OF TRANSCRIPTION LOOP SIZE, GENE SIZE AND TRANSCRIPTION LEVEL

#### 3.4.1 HIGHLY EXPRESSED GENES ARE EXPRESSED BIALLELICALLY

Monoallelic expression refers to the random expression of a gene from only one of the two alleles in a diploid cell and has been reported for > 10 % of autosomal mouse genes (Zwemer et al., 2012). Monoallelic expression is cell-type specific and leads to a high heterogeneity between different cells of the same tissue (Eckersley-Maslin and Spector, 2014). It was suggested as a mechanism to fine-tune gene dosage in order to control developmental regulatory pathways (Gendrel et al., 2014).

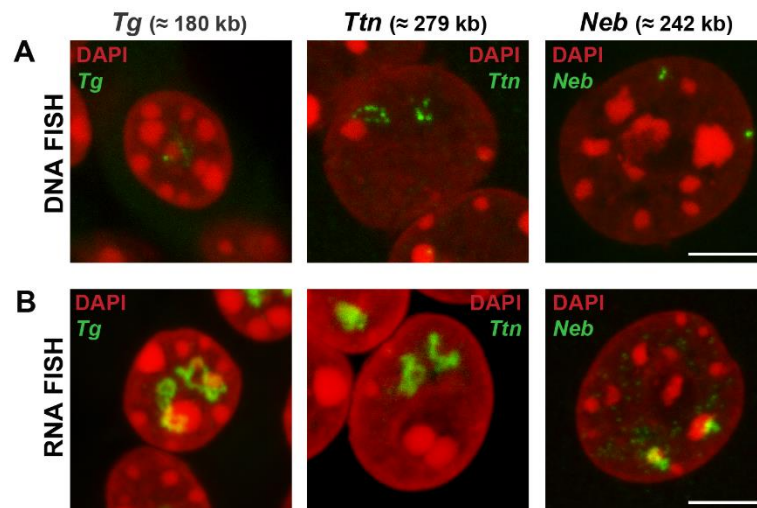
All five genes we investigated exhibit high and robust transcription levels and therefore could serve as convenient models to study random monoallelic expression by microscopy. Using RNA FISH on the respective gene, we scored nuclei with one *Tg* TL in thyrocytes or one *Ttn* TL in cultured myotubes or skeletal muscle. Only 5.4 % of myotube nuclei exhibited one active *Ttn* allele (TL). In tissues, the percentage of nuclei with only one active allele was even lower: 2 % exhibited one active *Ttn* allele in skeletal muscle and 3 % exhibited one active *Tg* allele in thyrocytes. *Neb*, *Myb11* and *Cald1* also consistently exhibited two TL signals when visualized by RNA FISH in respective tissues. Based on these observations, we conclude that highly upregulated genes are generally expressed biallelically.

#### 3.4.2 TRANSCRIPTION LOOP SIZE DOES NOT CORRESPOND TO GENE SIZE

Transcriptional bursting is considered to be a universal principle of gene expression (Chubb et al., 2006). It is characterized by a stochastic alternation between activation and inactivation of promoters, which results in the discontinuous production of mRNAs in bursts interrupted by pauses. The active periods of a promoter may range from minutes to hours leading to the production of a few to hundreds of transcripts during one burst (Nicolas et al., 2017). Furthermore, it was shown previously that during a transcription burst, RNA polymerases are loaded and move in groups, so-called convoys (Tantale et al., 2016). Therefore, one gene can feature multiple transcription bursts (RNA Pol II convoys) along its body. Assuming an elongation speed of 3.8 kb/min, the time needed for transcription of the entire *Ttn* gene is approx. 75 min, for *Neb* approx. 60 min, for *Tg* and *Cald1* approx. 45 min, and for *Myb 11* approx. 25 min. The presence of consistent biallelic RNA signals of the studied genes (see Results section 3.4.1) implies that in case of temporal silencing of one of the homologs, such pause must be less than 25 min.

A recent study has shown that the expression level of a gene is directly connected to the frequency as well as the duration of its transcription bursts (Larsson et al., 2019). Highly expressed genes are transcribed in long bursts, that is they are covered by long RNA Pol II convoys that are frequently initiated and interrupted by short pauses. Lowly expressed genes, in contrast, are transcribed in short bursts with long pauses (Larsson et al., 2019). Taking these considerations into account, we assumed that the TLs of the studied genes expressed on different levels, should have different bursting patterns reflected by their TL structure. Indeed, the patterns of DNA FISH signals strongly differ between the genes. Whereas the *Tg* gene (22,924 TPM) is highly

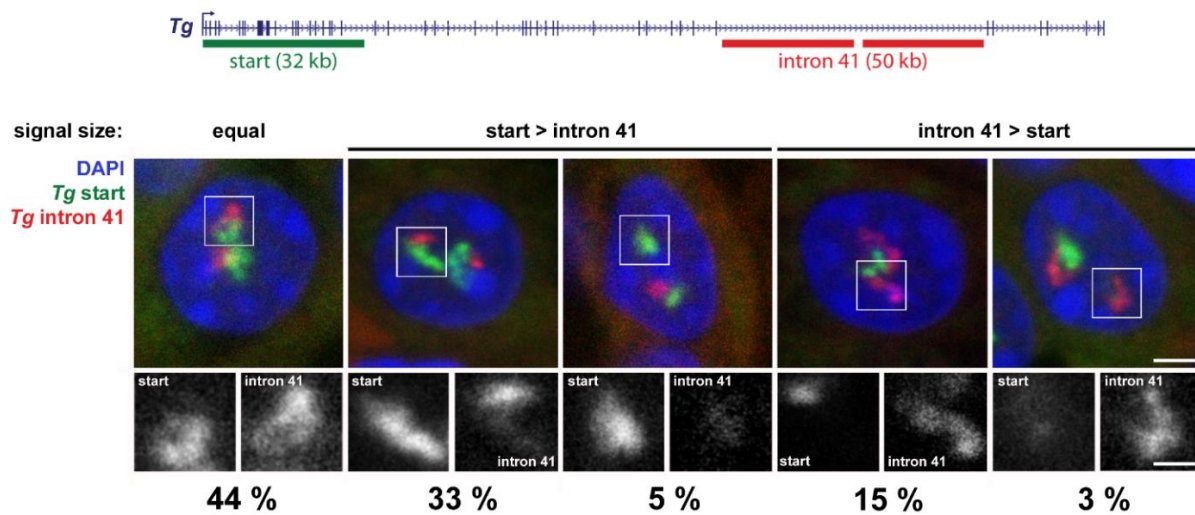
decondensed and its axis is hardly traceable, the *Ttn* gene (2,110 TPM) shows a beads-on-string structure with alternating decondensed and condensed regions (Figure 3.10 A). The gene body of *Neb* (980 TPM) exhibits a rather solid signal and exhibits only one or two resolvable gaps (Figure 3.10 A). We reason that the decondensed regions correspond to transcription bursts, while the condensed regions correspond to pauses between two bursts. For a given gene, the TL pattern, that is the number of DNA FISH foci, differs between different cells, indicating that each cell and even each allele exhibit their own bursting rhythm. Consistent with our reasoning, the size of RNA FISH signals ( $Tg > Ttn > Neb$ ) of the studied genes does not directly correlate with the length of the genes ( $Ttn: \approx 280 \text{ kb} > Neb: \approx 242 \text{ kb} > Tg: \approx 180 \text{ kb}$ ) but rather with their transcription level ( $Tg: 22,924 \text{ TPM} > Ttn: 2,100 \text{ TPM} > Neb: 980 \text{ TPM}$ ) (Figure 3.10 B). We therefore conclude that a gene's bursting kinetics is reflected in the structure of the formed TL.



**FIGURE 3.10: Transcription loop size does not correspond to gene length.** *Tg* transcription loops are generally bigger than the TLs formed by *Ttn* (279 kb) or *Neb* (242 kb), despite its smaller genomic size (180 kb). However, the *Tg* gene (22,924 TPM) exhibits a 10-fold higher transcription level than *Ttn* (2,110 TPM) and a 20-fold higher transcription level than *Neb* (980 TPM). **(A)** DNA FISH for the same genes in the respective cell types (thyrocytes for *Tg* and myotubes for *Ttn* and *Neb*) revealed a high decondensation of the *Tg* gene body, reflected by a hardly traceable FISH signal (left). In contrast, the gene bodies of *Ttn* and *Neb* exhibited condensed regions indicated by bright dots (middle and right, respectively), with the *Ttn* gene body consisting of five to eight such dots and the *Neb* gene body consisting of one or two dots. DNA was counterstained with DAPI (red). Scale bar 5  $\mu\text{m}$ . **(B)** RNA FISH for *Tg* in thyrocytes (left) and *Ttn* and *Neb* in myotubes (middle and right, respectively) revealed that *Tg* TLs are more expanded than those of *Ttn* and *Neb*. Scale bar 5  $\mu\text{m}$ .

## RESULTS

Transcriptional bursting of the *Tg* gene was further investigated by using a probe encompassing selective regions of the gene. RNA FISH with two differentially labeled probes encompassing the 5' end (32 kb) and intron 41 (50 kb) of the *Tg* gene (Figure 3.11) revealed varying patterns of bursting between homologues. 8 % of the *Tg* alleles did not have one of the signals, indicating that these TLs 5' or 3' regions were in between two bursts (Figure 3.11). Given that the length of the intron probe is approx. 50 kb, the data indicate that the pause between two bursts (the distance between two Pol II convoys) can be up to 50 kb. Additionally, analysis of the signal sizes of 87 *Tg* alleles showed that in 38 % of the alleles, the start signal deriving from a smaller genomic region is bigger than the intron signal, in 18 % of the alleles the intron signal is bigger, and in 44 % they are of equal size (Figure 3.11). The different sizes of RNA FISH signals and their absence in some cases suggest that alleles of the same gene exhibit different bursting patterns.



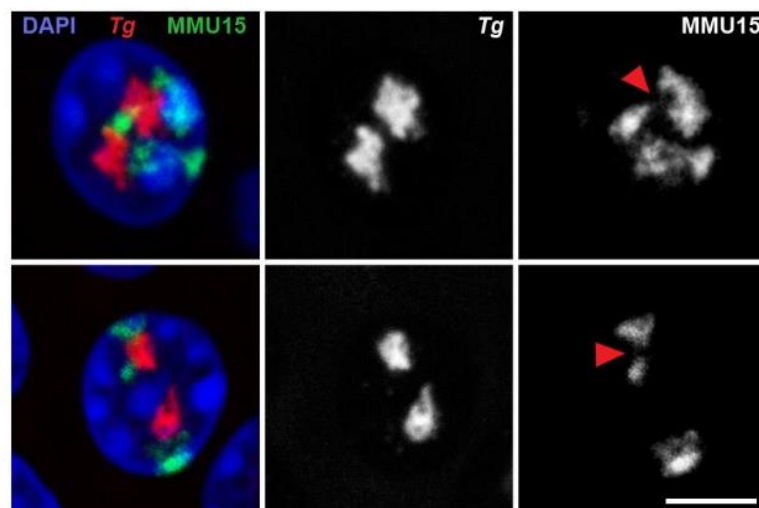
**FIGURE 3.11: Not all regions of *Tg* are transcribed at the same time.** RNA FISH with probes encompassing short ( $\approx 30$ –50 kb) regions of the *Tg* gene (top, beginning of *Tg* in green, end of *Tg* in red) reveals that not every part of the *Tg* gene is transcribed at any time. In 38 % of investigated alleles ( $n = 87$ ), the start signal is bigger than the intron signal, in 18 %, the intron signal is bigger and in 44 %, they are equally sized. 8 % of investigated alleles show an absence of signal of either of the two probes (5 %: absence of intron signal, 3 %: absence of start signal), indicating that this respective part of the gene is not covered by polymerases and therefore no RNA is transcribed. DNA was counterstained with DAPI (blue). Scale bars 2  $\mu\text{m}$  and 1  $\mu\text{m}$  (zoom).

### 3.5 TRANSCRIPTION LOOPS SHAPE THEIR IMMEDIATE CHROMATIN NEIGHBORHOOD

#### 3.5.1 *Tg* TRANSCRIPTION LOOPS ARE EXCLUDED FROM HARBORING CHROMOSOME TERRITORIES

Since chromosome territoriality has been considered a major hallmark of genome organization in the cell nucleus (see Introduction section 1.1.3), we investigated the spatial relationship between a highly transcribed gene forming a TL and its harboring chromosome. Using a paint for mouse chromosome 15 and a genomic probe for the *Tg* gene, we performed FISH simultaneously detecting DNA and RNA and observed that the transcribed *Tg* gene is consistently excluded from the harboring chromosome and forms its own “transcription territory” directly adjacent to the chromosome territory (CT) (Figure 3.12). Moreover, in 2 % of the alleles (n = 200), the *Tg* transcription loop splits the CT into two parts (Figure 3.12).

We also visualized *Ttn* and *Neb* TLs together with their harboring chromosome 2 in myotubes and found these genes residing either on the surface, or outside of the respective CT but not within the CT in the majority of the investigated alleles (90 % of *Ttn* alleles, n = 100; and 72 % of *Neb* alleles, n = 81).



**FIGURE 3.12: The *Tg* TL disrupts the territory of mouse chromosome 15.** Visualization of the *Tg* TL (red) together with the respective chromosome 15 (MMU15, green) in mouse thyrocytes. The TLs form their own “territory” next to the CT. In 2 % of alleles, the TL divides the CT into two parts (red arrowheads). DNA was counterstained with DAPI (blue). Scale bar 5  $\mu$ m.

### 3.5.2 Hi-C DATA CONFIRM EXCLUSION OF TRANSCRIPTION LOOPS FROM CHROMOSOMES

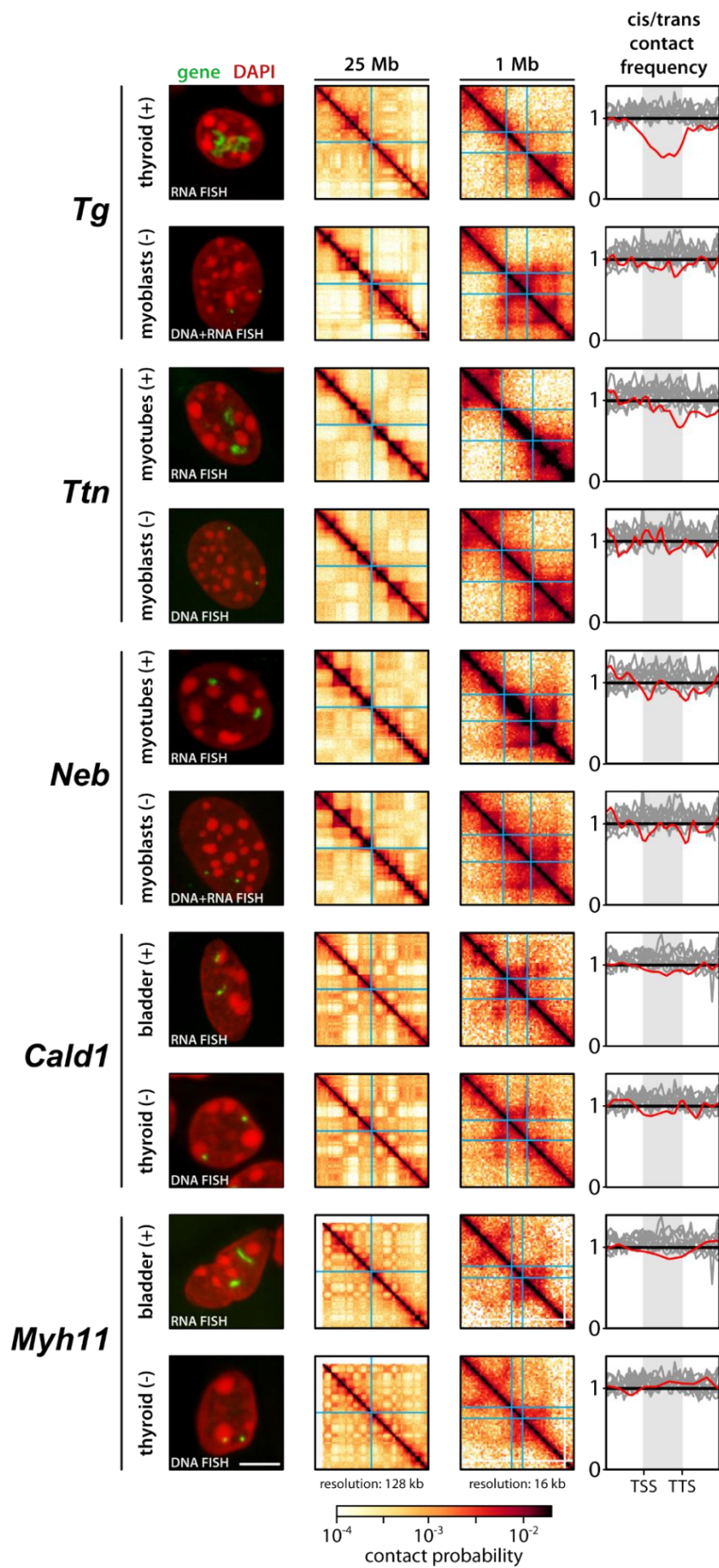
The relationship between the selected highly expressed genes and their harboring chromosomes was further quantified by the number of interactions between the respective gene and the rest of the genome. Four cell types were subjected to Hi-C: undifferentiated myoblasts, myotubes differentiated *in vitro*, thyrocytes (thyroid) and smooth muscle (bladder).

To comparably estimate the extent to which the TL forming genes are excluded from the respective CT, we calculated the ratio of *cis* (same chromosome) to *trans* (other chromosome) contacts along the gene body of the respective gene in the active and silent states. Analysis of these ratios revealed that genes exhibiting the largest TLs detected by FISH, such as *Tg* and *Ttn*, displayed a lower *cis/trans* contact frequency ratio than other long genes in the same cell type but with weaker expression (Figure 3.13). The decline in the *cis/trans* contact frequency ratio was especially prominent for the expressed *Tg* gene in thyrocytes (Figure 3.13). Notably, the dip in this case exhibits an asymmetrical shape with a deeper incline towards the TTS suggesting a more distinct exclusion from the CT of this part of the gene. Similarly, an asymmetric dip was also observed for expressed *Ttn* in myotubes. *Neb*, *Cald1* and *Myb11*, characterized by lower expression and visibly smaller TLs, have *cis/trans* ratios similar to other long lowly expressed genes (Figure 3.13).

Taken together, the observed transcription dependent decline in the *cis/trans* contact frequency ratio along the TL forming genes further suggests that long and highly expressed genes loop out of their respective CT.

**FIGURE 3.13: Hi-C data confirm the exclusion of TLs from the respective CT.** Left: FISH images depicting the shape of TL forming genes (green; *Tg*, *Ttn*, *Neb*, *Cald1* and *Myb11*) in the active (+) and silent (–) states. DNA was counterstained with DAPI (red). Scale bar 5  $\mu$ m. Middle: Hi-C maps and *cis/trans* contact ratios. 25 Mb Hi-C contact map and a 1 Mb zoom view with different resolutions showing contact frequencies in regions harboring TL forming genes in active and silent states. Blue cross hairs indicate the location of the respective gene. Surprisingly, the morphological features observed by microscopy are not reflected in Hi-C maps, where differences between “on” and “off” are only moderate and differ between the studied genes. Right: The ratio of *cis/trans* contact frequencies (number of *cis* reads/number of *trans* reads) along the gene bodies of *Tg*, *Ttn*, *Neb*, *Cald1* and *Myb11* was calculated based on Hi-C data in the active and silent states of the genes. Ratios along the genes of interest are depicted as red lines, dark gray lines correspond to long lowly expressed control genes. Gene body location is indicated by a light gray box. Ratios below 1 indicate a higher frequency of *trans* contacts in comparison to *cis* contacts and therefore suggest the exclusion of a region from the harboring CT. The *Tg* gene shows a pronounced *cis/trans* dip in thyroids; other highly expressed long genes show a moderate dip (*Ttn* in myotubes, *Myb11* in bladder). For comparability of *cis/trans* profiles, the x-coordinates are rescaled such that TSS and TTS of all genes in a panel align (gene length is marked by shaded area). Furthermore, to highlight potential dips in the gene body against longer range variations, *cis/trans* profiles are normalized to unity in the region outside the gene body (see Materials and Methods section 2.7 for details).

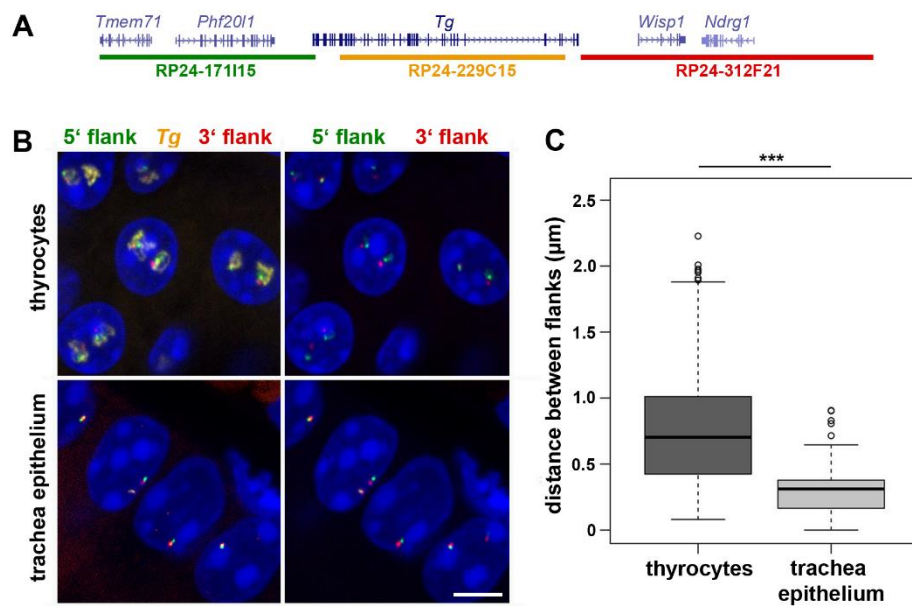




### 3.5.3 TRANSCRIPTION LOOPS ARE OPEN LOOPS WITH SEPARATED FLANKS

Earlier studies suggested that, in highly expressed genes, the promoter physically interacts with the transcription termination site (Hampsey et al., 2011). It is assumed that this interaction facilitates the recycling of polymerases since they directly engage in a new round of transcription immediately after being released at the 3' gene end (Shandilya and Roberts, 2012). Accordingly, transcriptional bursting has been attributed to the assembly and disassembly of such loops (Hebenstreit, 2013). Therefore, we were prompted to investigate whether in our selected highly expressed genes, the 5' and 3' ends reside in close proximity.

We conducted FISH detecting DNA and RNA simultaneously with probes hybridizing to the gene bodies of either *Tg* or *Tm* and to their respective flanking regions and measured 3D distances between the flanks using confocal stacks and a semi-automated script (see Materials and Methods sections 2.8.1 and 2.9.2). We found that distances (median) between *Tg* flanks (0.703  $\mu\text{m}$ ) are 2.3 times larger than in control cells (0.311  $\mu\text{m}$ ) not expressing the *Tg* gene (Figure 3.14).

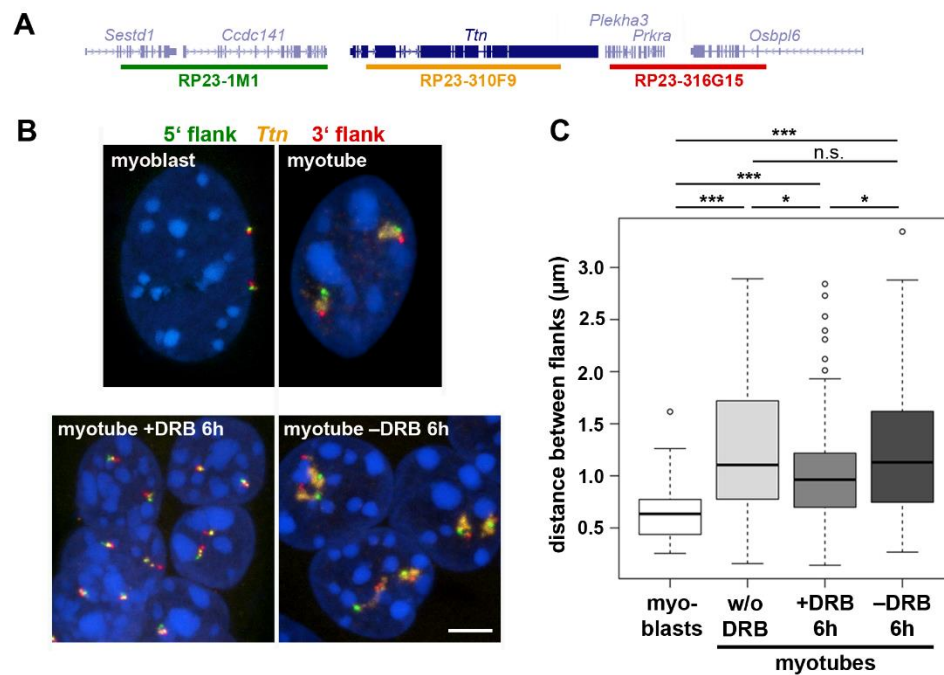


**FIGURE 3.14: The gene flanking regions of expressed *Tg* are spatially separated.** 85 % of alleles exhibit visibly separated 3' and 5' flanks of the expressed *Tg* gene in mouse thyrocytes. In cells not expressing *Tg* (trachea epithelial cells), the flanks separation is microscopically visible in only 20 % of alleles. **(A)** UCSC genome browser snapshot showing the position of *Tg* and neighboring genes. Color bars depict the location of BAC derived probes for *Tg* (yellow) and its 5' (green) and 3' (red) flanks. BAC clone identifiers are indicated below the bars. **(B)** Simultaneous detection of the *Tg* gene and its flanking regions (left) or flanking regions only (right) by FISH labeling both DNA and RNA in thyrocytes and control cells (trachea epithelial cells). DNA was counterstained with DAPI (blue). Scale bar 5  $\mu\text{m}$ . **(C)** Boxplot depicting the 3D distance between the flanking regions in the two tissues measured on confocal stacks. The inter-flank distances in thyrocytes ( $n = 203$ , median = 703 nm) expressing *Tg* is 2.3 times larger than in control cells ( $n = 180$ , median = 311 nm) not expressing *Tg*. \*\*\* $p$ -value  $< 0.001$ ; Wilcoxon rank sum test.



Similarly, median distances between *Ttn* flanks in myotubes (1.104  $\mu\text{m}$ ) expressing the gene were 1.7 times larger than in myoblasts (0.634  $\mu\text{m}$ ) in which the gene is silent (Figure 3.15).

Transcription inhibition with DRB shortened the distance between the flanks to 0.963  $\mu\text{m}$ , whereas transcription restoration after DRB removal caused convergence of the *Ttn* flanks to a median distance of 1.130  $\mu\text{m}$ , a value not significantly different from the distance measured in untreated myotubes (Figure 3.15). These results indicate that (i) transcription actively shapes the chromatin environment at the respective locus, and (ii) gene flanking regions are driven to diverge upon transcriptional induction.



**FIGURE 3.15: The separation of gene flanking regions is highly dynamic.** Measurements of gene flank distances in myoblasts, untreated myotubes as well as myotubes after reversible transcription inhibition via DRB, and myotubes with restored transcription after DRB removal, showed that the *Ttn* locus and its flanking regions exhibit dynamic reshaping depending on the gene's transcriptional state. **(A)** UCSC genome browser snapshot showing the position of *Ttn* and neighboring genes. Color bars depict the location of BAC derived probes for *Ttn* (yellow) and its 5' (green) and 3' (red) flanks. BAC clone identifiers are indicated below the bars. **(B)** Simultaneous detection of the *Ttn* gene and its flanking regions by FISH labeling both DNA and RNA in myoblasts, untreated myotubes, myotubes treated with DRB for 6h, and myotubes 6h after DRB removal. DNA was counterstained with DAPI (blue). Scale bar 5  $\mu\text{m}$ . **(C)** Boxplot depicting the 3D distances between the flanking regions of *Ttn* in the 4 different conditions measured on confocal stacks. The inter-flank distances in myotubes (w/o DRB;  $n = 116$ , median = 1104 nm) expressing *Ttn* was 1.7 times larger than in myoblasts (blasts;  $n = 62$ , median = 634 nm), i.e. cells not expressing the gene. Upon transcription inhibition with DRB, the gene flanks converged (+DRB 6h;  $n = 113$ , median = 963 nm), after DRB removal and transcription restoration, the gene flanks diverged again (-DRB 6h;  $n = 87$ , median = 1130 nm). \*\*\* $p$ -value < 0.001, \* $p$ -value < 0.05, n.s. not significant; Wilcoxon rank sum test.

### 3.6 DECIPHERING THE CAUSE OF TRANSCRIPTION LOOP FORMATION

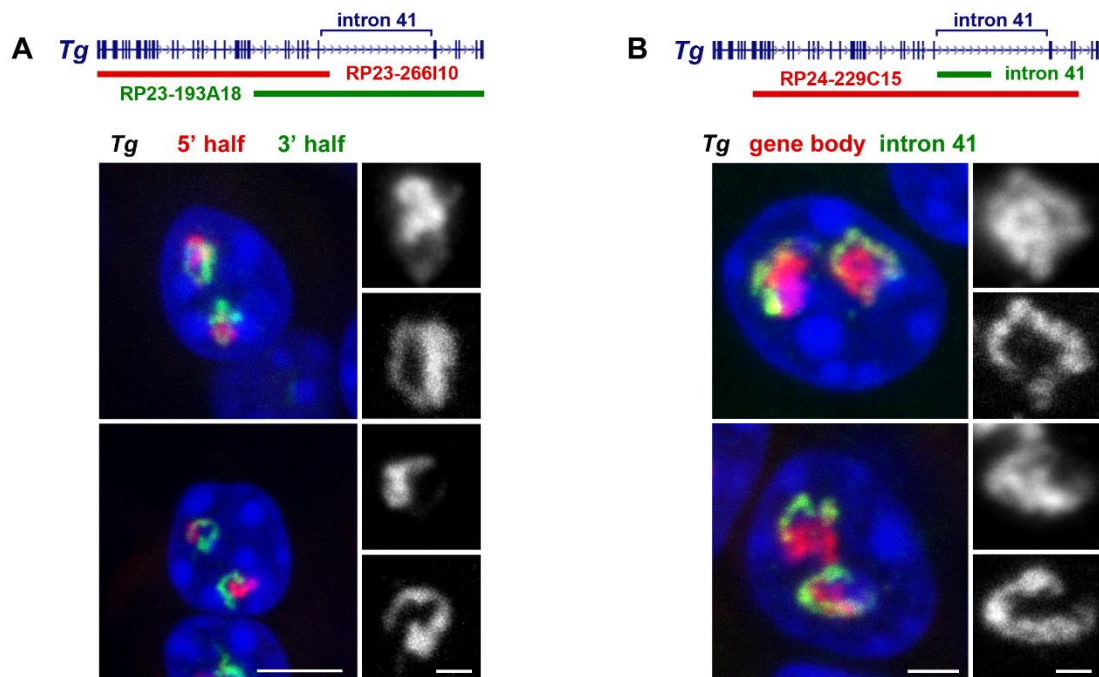
#### 3.6.1 ARE TRANSCRIPTION LOOPS STIFF STRUCTURES?

Collectively, the data presented in chapter 3.5 concerning the exclusion of TLs from harboring chromosomes and the separation of gene flanking regions point towards an intrinsic stiffness of TL structures. We hypothesize that such rigidity of the highly expressed genes could be caused by the dense decoration of their axis with RNA Pol II carrying nascent RNA transcripts. Nascent RNAs are known to bind hundreds of proteins involved in splicing, mRNA export, capping, poly adenylation, etc. (Muller-McNicoll and Neugebauer, 2013) and therefore form nascent ribonucleoprotein particles (nRNPs). In case of a long nascent RNA, the nRNPs decorating the gene axis therefore are quite voluminous exceeding the volume of nucleosomes and even polymerase complexes. When a gene is highly expressed, voluminous RNPs, attached to polymerases traveling in convoys, are densely spaced along the gene axis and thus render the gene axis less flexible but stiff. This in turn causes looping of a gene out of its harboring chromosome territory and protruding into the nuclear interior.

To test this hypothesis, we chose three different strategies. First, we investigated via FISH if the region of a TL corresponding to a long intron is markedly more expanded than a region corresponding to alternating short introns. Second, we employed EM to visualize nRNPs formed on the *Tm* gene. Third, we used polymer models of TLs with varying parameters in order to explore which of the models recapitulates our biological observations.

#### 3.6.2 THE LONGEST INTRON OF THE *Tg* GENE IS STRONGLY EXPANDED

In FISH experiments differentially labeling 5' and 3' halves of the *Tg* gene, the two halves are often stretched to different extents with a more expanded 3' half (Figure 3.16 A). This 3' half of the gene includes the long intron 41 measuring 53.8 kb (see also Results section 3.2.5). During transcription, nascent RNA (nRNA) grows from 7.3 kb to 61.1 kb across this intron. Since we demonstrated that the intron does not undergo microscopically detectable recursive splicing (see Results section 3.2.5), we expected that in case of a transcription burst including the entire intron, it should be expanded to a greater extent than the rest of the *Tg* TL. To estimate expansion of the intron, we designed a probe consisting of closely spaced fragments of approx. 200 bp obtained by PCR (see Material and Methods section 2.4.4). The probe encompasses the first half of the *Tg* intron 41 and therefore hybridizes to all nRNAs of the intron 41 before its excision. RNA FISH with a combination of the intron probe and a genomic probe encompassing most of the *Tg* gene showed that in the majority of the alleles, the part of the *Tg* TL corresponding to the large intron 41 is much more extended than the rest of the gene (Figure 3.16 B). Based on these observations, we concluded that (i) different TL regions are coiled to various degrees and (ii) the degree of coiling is dependent on the size of the associated transcripts: regions corresponding to short transcripts are coiled stronger than regions with longer transcripts.



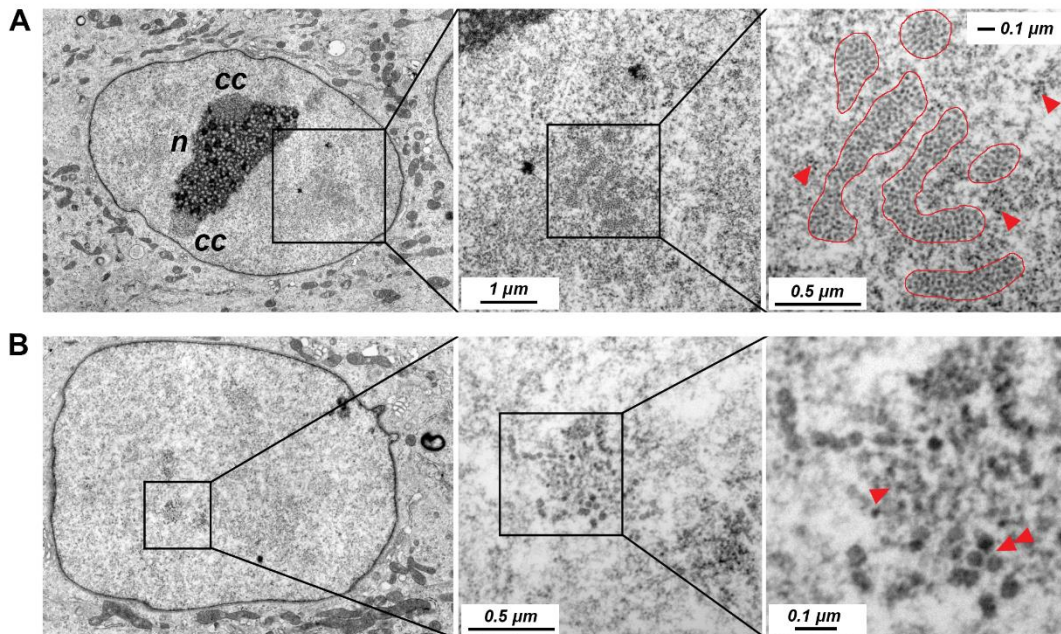
**FIGURE 3.16: The large *Tg* intron substantially contributes to TL expansion.** Intron 41, located in the 3' half of the *Tg* gene, measures 53.8 kb and exhibits greater expansion than the first gene half as shown by RNA FISH with BAC- and PCR derived probes. **(A)** Two examples of thyrocyte nuclei with TLs differentially labeled with two BAC probes for the first (red) and the second (green) half of the *Tg* gene. The signal deriving from the second half of the gene was expanded whereas the first half was rather coiled. Probe location is indicated by a scheme of *Tg* and red and green bars above the image panel. BAC identifiers are indicated next to bars. Vertical bars correspond to exons. DNA was counterstained with DAPI (blue). Scale bars 5  $\mu\text{m}$  and 1  $\mu\text{m}$  (zoom). **(B)** Two examples of thyrocyte nuclei with TLs labeled with a BAC probe encompassing almost the whole *Tg* gene (red) and a PCR derived probe specific for the first half of the 53.8 kb intron 41 (green). The signal deriving from the intronic probe was expanded comparable to the signal of the second half of the gene in (A). Probe location is indicated by a scheme of *Tg* and red and green bars above the image panel. BAC identifiers are indicated next to bars. Vertical bars correspond to exons. DNA was counterstained with DAPI (blue). Scale bars 2  $\mu\text{m}$  and 1  $\mu\text{m}$  (zoom).

### 3.6.3 ELECTRON MICROSCOPY REVEALS UNIQUE ACCUMULATIONS OF GRANULES IN MYOTUBE NUCLEI

*Electron microscopy was carried out in collaboration with Dr. Mandy Börmel and Dr. Yannick Schwab at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany.*

A literature search revealed that unambiguous and reliable identification of nRNPs so far was performed on polytene chromosomes, active at the larval stage in salivary glands of midges from the genus *Chironomus* (Mehlin and Danesholt, 1993; Olins et al., 1992; Stevens and Swift, 1966). The three highly upregulated genes on *Chironomus* chromosome IV are strongly decondensed and form huge polytene puffs, so called Balbiani rings (Bjork and Wieslander, 2015). The almost intronless genes measure about 35–40 kb and are decorated by nRNPs progressively growing in size, reaching a diameter of 50 nm at the end of the genes (Bjork and Wieslander, 2015).

For visualization of RNPs, we chose the *Ttn* gene because its mRNA measures  $\approx 103$  kb and therefore RNPs formed towards the middle or the end of the *Ttn* gene are likely detectable with electron microscopy. We investigated the ultrastructure of nuclei from *in vitro* differentiated myotubes and found conspicuous accumulations of densely packed granules of varying sizes from approx. 30 nm to approx. 50 nm in diameter (Figure 3.17 A). Smaller granules were packed in dense accumulations, whereas larger granules were packed more loosely (Figure 3.17 B). 40 % of examined myotube nuclei exhibited granular accumulations of approx. 30 nm, 30 % of nuclei exhibited larger granules of approx. 50 nm ( $n = 152$ ).



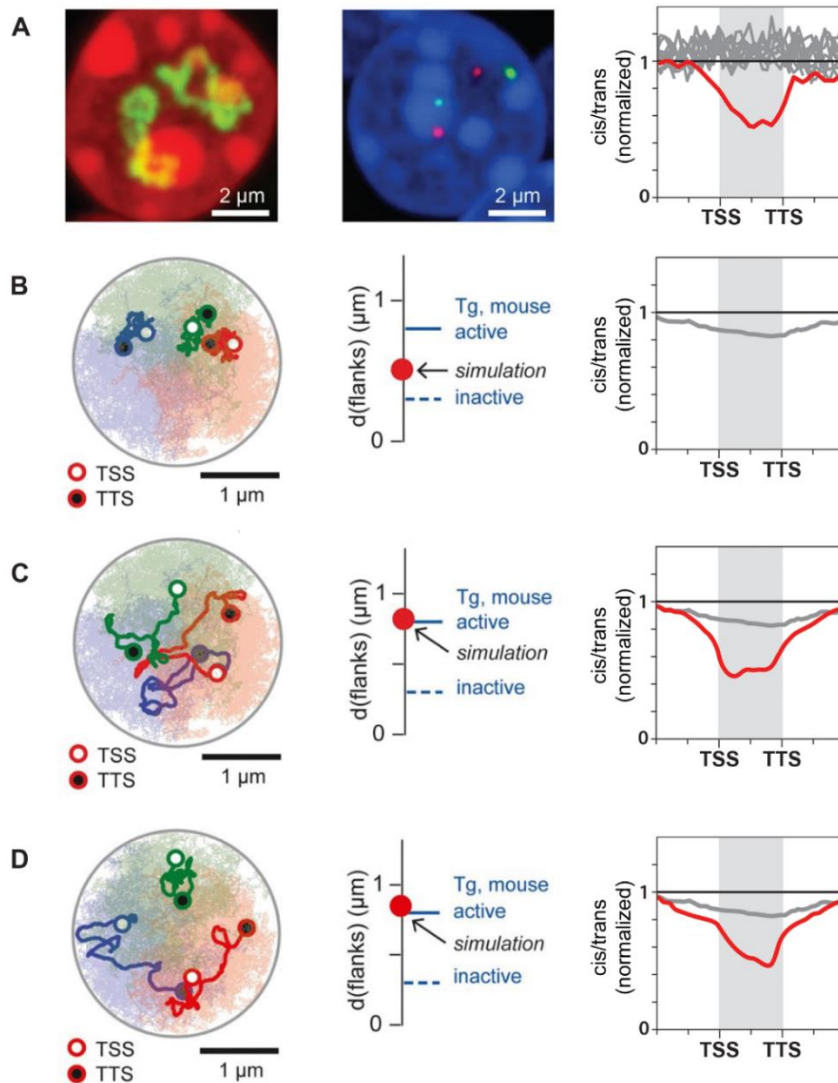
**FIGURE 3.17: Myotube nuclei exhibit granular accumulations.** Transmission electron microscopy revealed accumulations of 30 (A) to 50 (B) nm granules in myotube nuclei that typically reside within nuclear speckles as indicated by small granules with fibrillar connections (red arrowheads). **(A)** Myotube nucleus depicted in increasing magnifications (left to right) highlighting dense granular accumulations in euchromatin (right, outlined in red). Each individual granule measured approx. 30 nm. *n* = nucleolus, *cc* = chromocenter. **(B)** Myotube nucleus depicted in increasing magnifications (left to right) highlighting dense granular accumulations with sizes up to 50 nm (red doubled arrowheads) found in 30 % of myotube nuclei ( $n = 82$ ).

### 3.6.4 POLYMER MODELING CONFIRMS TRANSCRIPTION LOOP STIFFNESS

*Polymer modeling was carried out in collaboration with Dr. Johannes Nübler and Prof. Dr. Leonid Mirny at the Institute for Medical Engineering and Science and Physics at the Massachusetts Institute of Technology (MIT) in Cambridge, USA.*

To investigate whether TL stiffness indeed is the main determinant of TL formation, we sought to model genes with various physical properties and compared the resulting models with biological observations on the *Tg* gene. Six chromosomes were modeled as 50 Mb polymers with a monomer size of 1 kb. Simulated mitotic-like conformations were allowed to expand within a spherical nucleus to achieve chromosomal territoriality. On each simulated chromosome, a region of 100 kb was assigned as gene of interest. By changing different parameters of such a gene, we explored at which conditions biological observations on the *Tg* TLs were recapitulated – the visual appearance of the TLs, including loop size and degree of coiling, the distances between the gene flanking regions and the ratio of *cis/trans* contacts along the gene (Figure 3.18 A). For instance, a mere (3-fold) increase in the gene length as result of a partial nucleosome loss as it has been found for highly expressed genes (Kulaeva et al., 2010), did neither lead to the formation of TLs with separate flanks nor to changes in *cis/trans* ratio of contact frequencies calculated from a simulated Hi-C (Figure 3.18 B). Only models employing both a 3-fold gene length extension and a 12-fold increase in gene stiffness successfully recapitulated all three parameters – the morphology of *Tg* TLs, inter-flank distances and the decline in the ratio of *cis/trans* contact frequencies (Figure 3.18 C). Models implicating an increase in gene stiffness from the 5' towards 3' gene end mimicking tapered gene stiffness due to growing nRNPs performed even better. Analysis of simulated Hi-C of these models exhibited asymmetrical *cis/trans* ratio curves with a steeper incline towards the 3' end (Figure 3.18 D), similar to the curves derived from biological experiments (compare with *Tg* and *Tm* in Results section 3.5.2, Figure 3.13).





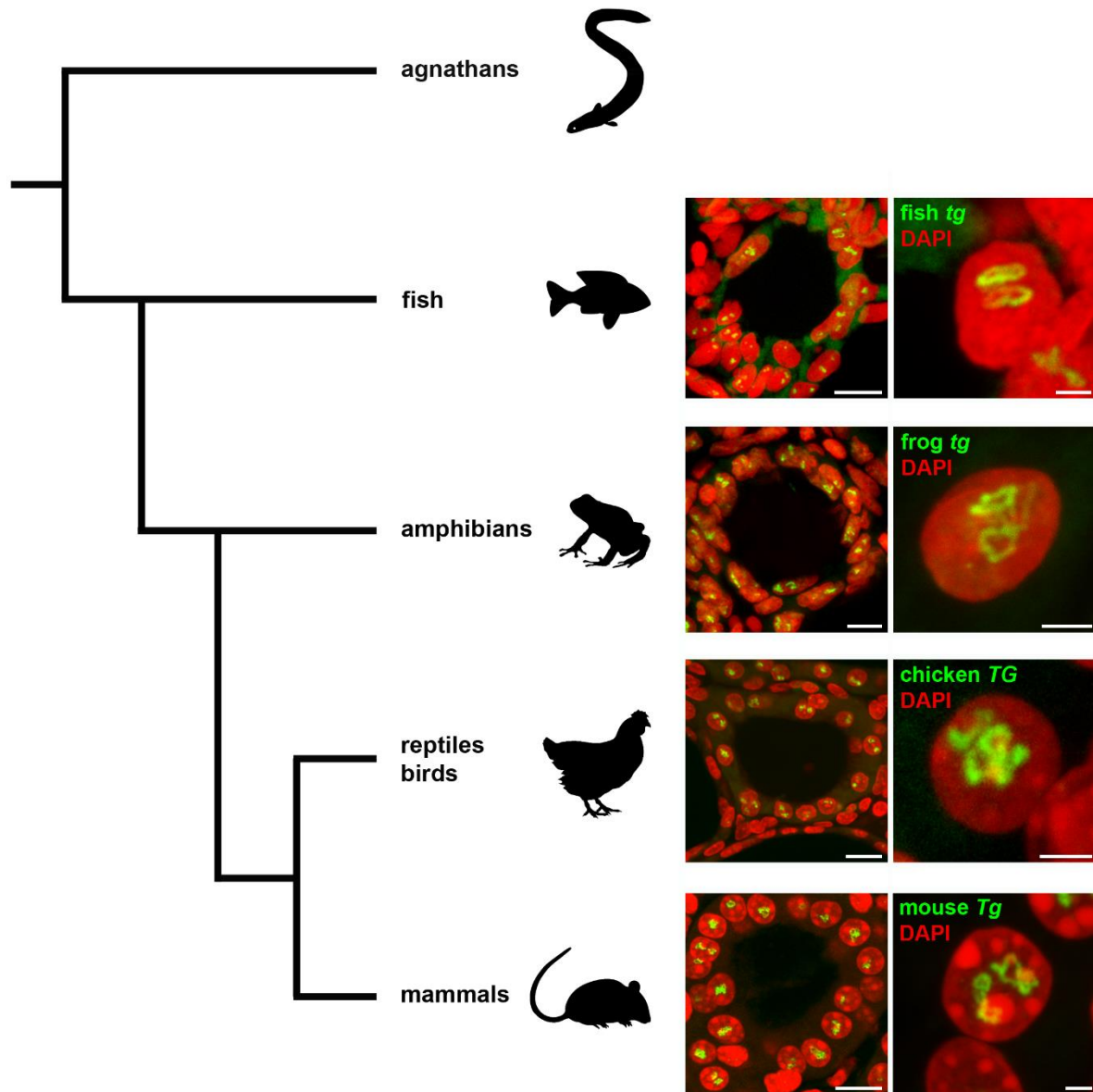
**FIGURE 3.18: Polymer models implicating gene stiffness recapitulate TL formation.** Polymer models confirm that a mere increase in gene length does not lead to TL formation. Only an increase in both gene length and gene stiffness recapitulates biological observations on the *Tg* TL. **(A)** Biological observations on *Tg* in thyrocytes used to determine model performance. From left to right: visual appearance of the *Tg* TL (green, DNA was counterstained with DAPI (red)), inter-flank distances of 776 nm (mean), HiC map and a plot depicting the respective *cis/trans* contact frequency ratio along the gene (red line; long, lowly expressed genes serve as comparison (grey lines)). Scale bars 2  $\mu$ m. **(B) – (D)** Polymer models with varying parameters. From left to right: Visual appearance of the genes of interest (red, blue and green) within modeled nuclei, simulated inter-flank distances (red dot, actual measured distances for active and inactive genes are indicated by solid and dashed blue lines, respectively), simulated HiC maps and a plot depicting the respective *cis/trans* contact frequency ratio along the gene (grey background) and its immediate surrounding. All models implicate 6 chromosomes of 50 Mb each with regions of 100 kb assigned as genes of interest. **(B)** In first model, the gene length was increased 3-fold. This mere increase in gene length yielded shorter inter-flank distances than measured and did not lead to a decrease in the *cis/trans* contact frequency ratio. **(C)** Same model as in (B), but with a 12-fold increase in gene stiffness. This model recapitulates both, the measured inter-flank distances and the decrease in the *cis/trans* contact frequency ratio (red line). *Cis/trans* contact frequency ratio from (B) is shown in dark grey as comparison. **(D)** Same model as in (B) but with a tapered stiffness along the gene body increasing towards the TTS. This model also recapitulates both the measured inter-flank distances and the decrease in the *cis/trans* contact frequency ratio, but performs even better. Note that the dip in *cis/trans* contact frequency ratio along the gene body (red line) is asymmetrical with a slightly deeper decrease towards the TTS that is also seen in (A). *Cis/trans* contact frequency ratio from (B) is shown in dark grey as comparison.

### 3.7 TAKING A CLOSER LOOK AT THE EXTRAORDINARY THYROGLOBULIN GENE

The mouse *Tg* gene is especially unique in terms of its remarkably high expression level combined with its length. After characterizing *Tg* as a TL forming gene, we aimed at further exploiting this extraordinary gene to investigate TL formation with regard to evolutionary aspects, systemic influences and regulatory pathways beyond direct transcriptional output. Therefore, experiments focusing on the *Tg* gene are illustrated and discussed in this separate chapter.

#### 3.7.1 *TG* FORMS TRANSCRIPTION LOOPS IN OTHER VERTEBRATE SPECIES

Given the importance of T3 and T4 hormones for vertebrate physiology and development (Iwen et al., 2013; Little, 2006a, b), we were prompted to investigate whether the *TG* gene is highly upregulated not only in mammals, but also in other vertebrate groups. To this end, we conducted RNA FISH experiments on thyrocytes of other vertebrate species. We prepared cryosections of chicken (*Gallus gallus domesticus*) and frog (*Xenopus tropicalis*) thyroid glands, as well as heads of fish (*Danio rerio*). Fish thyrocytes were identified based on green fluorescence deriving from eGFP expression under the TG promoter (see Materials and Methods section 2.2.2). We found that in all of these species, similar to mouse thyrocytes, both alleles of the *Tg* gene form transcription loops (Figure 3.19), suggesting that *Tg* is highly upregulated in all vertebrate groups.



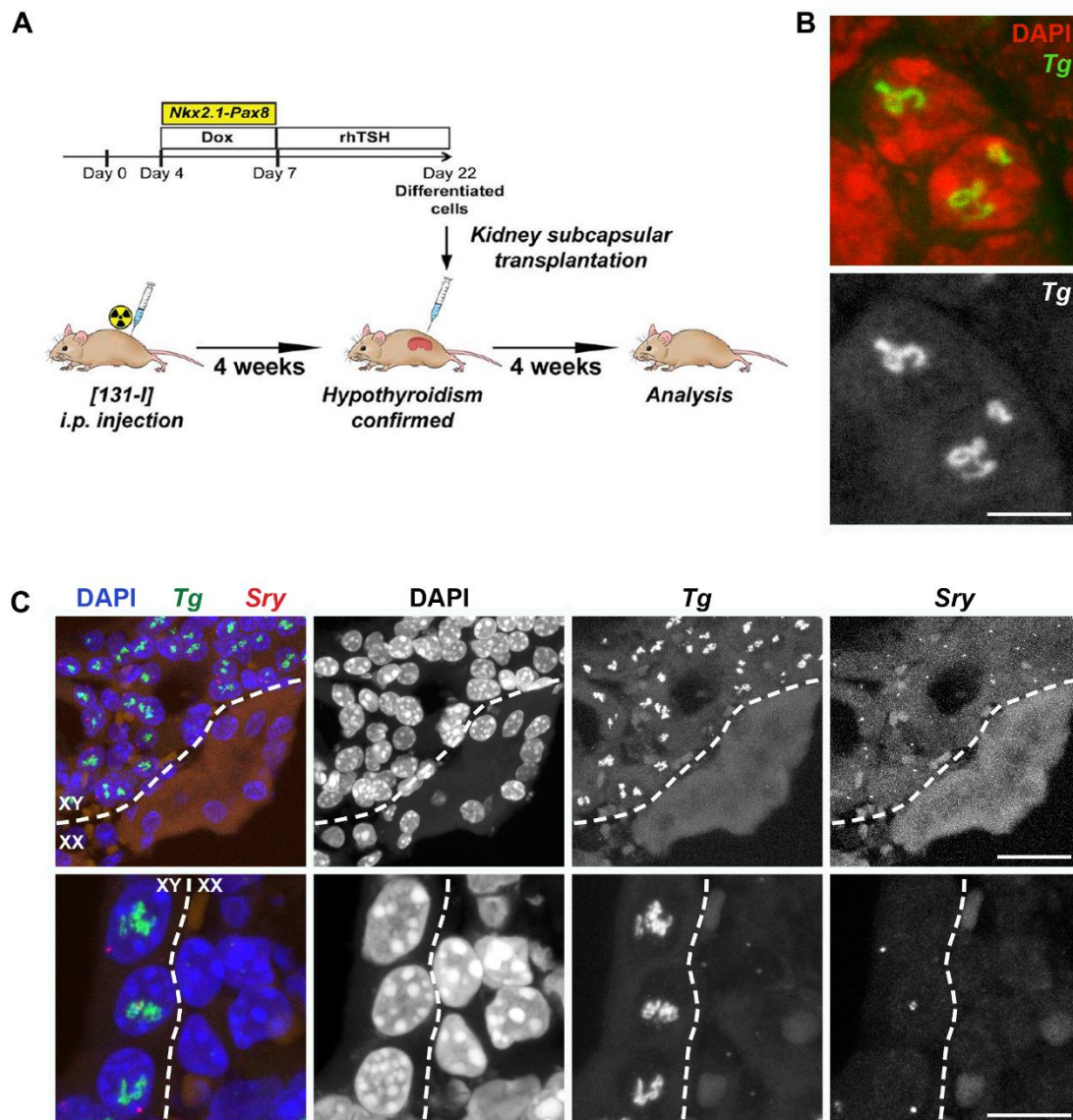
**FIGURE 3.19: The *Tg* gene forms transcription loops in all other vertebrate groups.** TL formation in other vertebrate species was assessed by RNA FISH with genomic BAC probes in thyroid tissue of different vertebrate groups (icons). From top to bottom: Zebrafish (*Danio rerio*) *tg* ( $\approx 68$  kb), frog (*Xenopus tropicalis*) *tg* ( $\approx 122$  kb) and chicken (*Gallus gallus domesticus*) *TG* ( $\approx 139$  kb) all show TL comparable to those of mouse (*Mus musculus*, *Tg*  $\approx 180$  kb) thyrocytes. Phylogenetic tree is shown for reference. Scale bars  $10\ \mu\text{m}$  (overview) and  $2\ \mu\text{m}$  (zoom).



### 3.7.2 *Tg* FORMS TRANSCRIPTION LOOPS IN *IN VITRO* GROWN AND IN GRAFTED THYROID FOLLICLES

In 2012, researches succeeded in generating functional thyroid tissue from mouse embryonic stem cells (mESCs) by overexpressing the thyroid specific transcription factors NKX2.1 and PAX8 in combination with thyrotropin (TSH) treatment *in vitro* (Antonica et al., 2012). The obtained tissue was grafted into the kidneys of athyroid mice, rescued hormone plasma levels, and also led to recovery of athyroidism derived symptoms (Antonica et al., 2012) (Figure 3.20 A). Since this thyroid tissue appears to be fully functional, we reasoned that also the transcription rate of *Tg* should be as high as in endogenous mouse tissue.

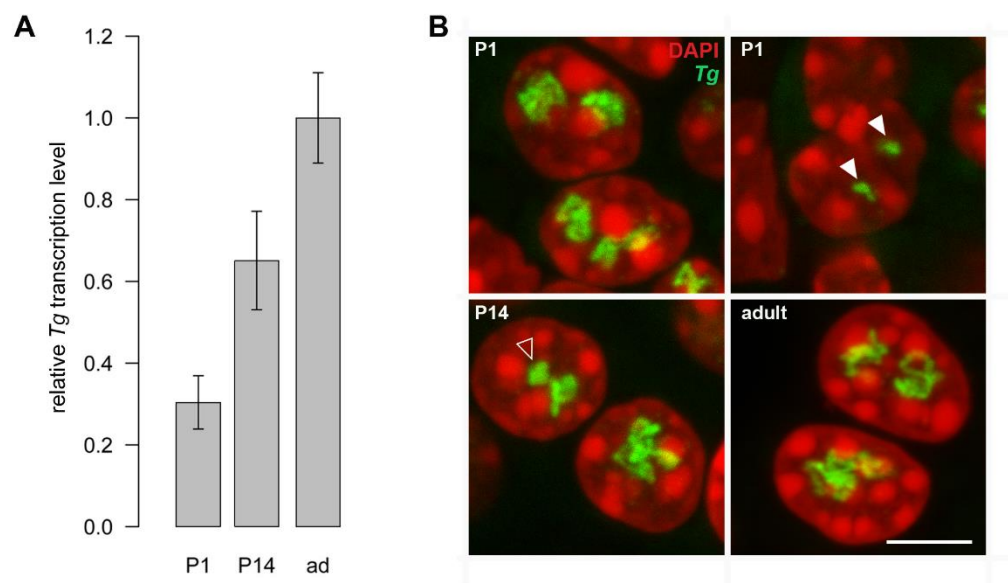
We conducted FISH detecting both DNA and RNA with a genomic BAC probe for *Tg* on cyrosections of *in vitro* grown thyroid follicles as well as grafted kidney tissue provided by Dr. Sabine Costagliola (Université libre de Bruxelles, Brussels, Belgium), and found TLs in follicles grown in matrigel *in vitro* (Figure 3.20 B). However, the structure and fine morphology of these TLs were hardly comparable to *Tg* TLs in endogenous mouse thyroid tissue due to the compromised morphology of thyrocyte nuclei after FISH on matrigel embedded follicles (Figure 3.20 B). Therefore, we also performed FISH on cryosections derived from the kidney of a female mouse with implanted cultured follicles (Antonica et al., 2012). To identify the grafted male mouse tissue, we used a BAC spanning sex-determining region Y (*Sry*) on the Y chromosome. In this case, FISH revealed *Tg* TLs comparable in shape and size to those of endogenous mouse thyroid tissue (Figure 3.20 C). This data indicate that *Tg* TLs are formed in functional thyrocytes irrespective of their origin and that tissue functionality coincides with *Tg* TL formation.



**FIGURE 3.20: *Tg* forms a TL in grafted thyroid follicles generated from mESCs *in vitro*.** mESCs transiently expressing the TFs NKX2.1 and PAX8 were differentiated *in vitro* to form functional thyroid tissue. When transplanted into hypothyroid mice, these tissues rescue blood thyroid hormone levels (Antonica et al., 2012). **(A)** Schematic depiction of the workflow to generate thyroid follicles from mESCs *in vitro* and to test tissue functionality. TFs Nkx2.1 and Pax8 are overexpressed under the control of a doxycycline inducible promoter. Then, cells are treated with recombinant human TSH for 15 days. After a total of 22 days, functional thyroid follicles have formed. These follicles are grafted under the renal capsule of hypothyroid mice previously subjected to radioiodine ablation. From (Antonica et al., 2012). **(B)** FISH detecting both DNA and RNA with a BAC probe for *Tg* in *in vitro* grown thyroid follicles before transplantation. *Tg* (green) also forms transcription loops in these follicles comparable to the *Tg* TL found in endogenous mouse tissue. DNA was counterstained with DAPI (red). Scale bar 5  $\mu$ m. **(C)** FISH detecting both DNA and RNA with a BAC probe for *Tg* in *in vitro* grown thyroid follicles after subcapsular renal transplantation. The transplanted tissue originates from male mESCs and is therefore distinguishable from the surrounding female host tissue by FISH using a probe for the *Sry* (sex-determining region Y) gene located on the Y chromosome (red). The grafted thyrocytes exhibit giant TLs (green) similar to thyrocytes of the native thyroid gland. DNA was counterstained with DAPI (blue). Scale bars 25  $\mu$ m (upper panel) and 10  $\mu$ m (lower panel).

### 3.7.3 *Tg* IS HIGHLY UPREGULATED AND FORMS TRANSCRIPTION LOOPS UPON BIRTH

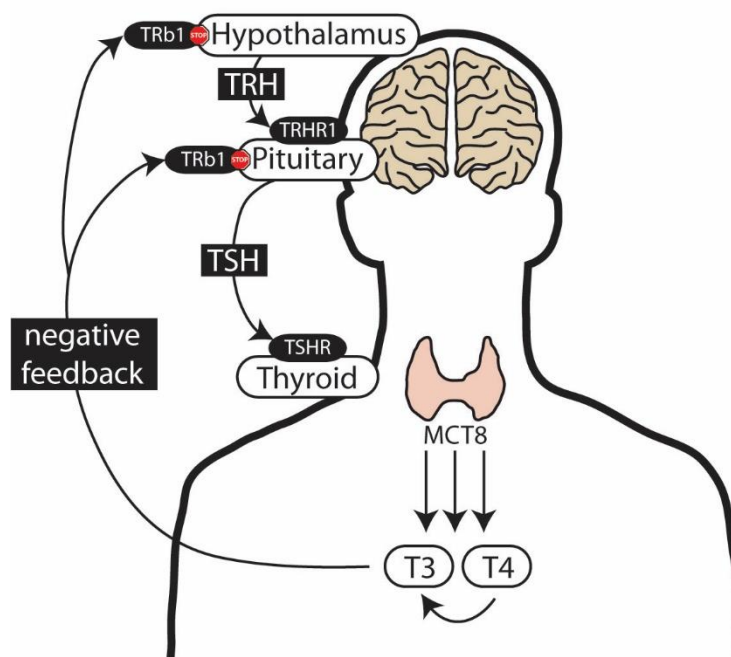
Next, we asked when the formation of *Tg* TLs is initiated during development and whether TL size and transcription level are concurrent during the onset of developmental *Tg* upregulation. Therefore, we first performed qPCR on thyroid glands from different developmental stages (P1, P14 and adult) to determine *Tg* transcription levels. We showed that *Tg* expression increases during development, reaching a 3.3-fold elevation at the adult stage compared to P1 (Figure 3.21 A). Interestingly, the transcription level was also reflected by the size of the TLs in the respective developmental stages. Even if most of the alleles exhibited huge TLs comparable to the TLs found in adult animals, some P1 thyrocytes had small loops coinciding with a lower transcription level (Figure 3.21 B). P14 thyrocytes, exhibiting a medium level of *Tg* expression, also harbored small TLs, however, these loops were larger than the ones found at the P1 stage (Figure 3.21 B). In contrast, thyrocytes of adult animals featuring extremely high *Tg* expression, stably displayed large TLs (Figure 3.21 B). Therefore, we conclude that *Tg* TL loop formation is established early in development, and that TL size correlates with developmental stage due to a gradually established *Tg* expression level.



**FIGURE 3.21: The *Tg* gene forms TLs early in development.** (A) *Tg* transcript levels gradually increase throughout development from P1 to the adult stage as shown by qPCR (relative transcription levels: P1: 0.30; P14: 0.65; ad: 1). ad = adult. Bars represent standard deviation. (B) RNA FISH with a genomic BAC probe for *Tg* reveals huge TLs (green) in all of the tested developmental stages (P1, P14, adult). However, some undeveloped TLs are found in thyrocytes from P1 (filled arrowheads) and P14 (empty arrowheads) mice. These small undeveloped loops correspond well to the lower expression level of *Tg* in earlier stages as shown in (A). DNA was counterstained with DAPI (red). Scale bar 5  $\mu$ m.

### 3.7.4 *TG* IS HIGHLY EXPRESSED AND FORMS TRANSCRIPTION LOOPS IRRESPECTIVE OF THYROID HORMONE LEVELS

Thyroid hormones positively influence cellular metabolism and therefore are important regulators of development (Brent, 2012; Iwen et al., 2013; Little, 2006a, b; Oetting and Yen, 2007). The production of these hormones is regulated via a series of interconnected feedback loops on the hypothalamus-pituitary-thyroid axis (DiStefano, 1973; Hoermann et al., 2015; Ortega-Carvalho et al., 2016). A simplified scheme depicting basic pathways of thyroid hormone regulation is provided in Figure 3.22: in case of low blood hormone levels, the hypothalamus secretes thyrotropin releasing hormone (TRH). TRH binds to thyrotropin releasing hormone receptors (TRHRs) on the surface of thyrotropes in the pituitary gland. These cells in turn secrete thyroid stimulating hormone (TSH) which stimulates the thyroid gland to produce thyroid hormones T3 and T4. Thyroid hormones are secreted into the blood stream via the mono-carboxylate transporter 8 (MCT8). When thyroid hormone levels in the blood exceed the necessary level, the hormones bind to thyroid hormone receptor beta (TRb1) in the hypothalamus and the pituitary gland signaling to reduce TRH and TSH hormone production.



**FIGURE 3.22: Schematic representation of the regulation of thyroid hormone levels.** Thyroid hormone production involves three major players, the hypothalamus, the pituitary gland as well as the thyroid itself, and is controlled via a negative feedback loop. The hypothalamus secretes thyrotropin releasing hormone (TRH) which binds to TRH receptors (TRHR1) in the pituitary gland. The pituitary in turn secretes thyroid stimulating hormone (TSH) which binds to TSH receptors in the thyroid gland and signals the gland to produce triiodothyronine (T3) and thyroxine (T4) and release them through MCT8 transporters. T4 is converted to the active form T3 by deiodinases intercellularly. If thyroid hormone blood levels reach a certain threshold, they bind to thyroid hormone receptor b1 (TRb1) in hypothalamus and pituitary, and thus signal to lower TRH and TSH release in order to eventually reduce T3 and T4 levels in the blood stream.

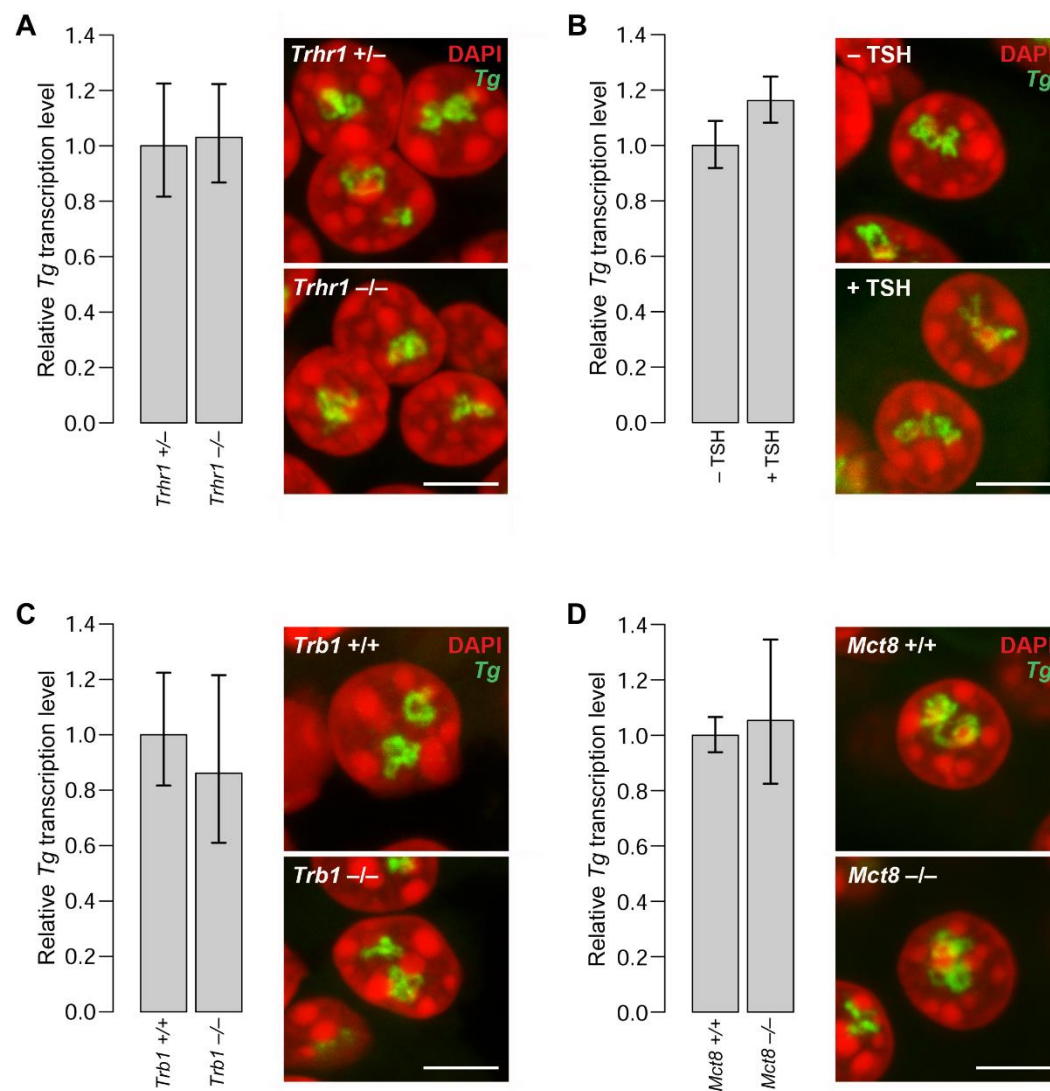
Misregulation of the gland's hormonal output causes hypo- or hyperthyroidism, manifested through insufficient or excessive hormone production, respectively (Little, 2006a, b). In order to test whether the misregulation of hormone production influences the formation of TLs, the topology of the *Tg* gene in mice with induced hypo- or hyperthyroidism was investigated.

Thyroid tissues from hypo- and hyperthyroid mice provided by Prof. Dr. Heike Heuer (Universitätsklinikum Essen, Klinik für Endokrinologie und Stoffwechselerkrankungen, Essen, Germany) were subjected to qPCR and RNA FISH. Hypothyroid mice were generated by deleting both alleles of *Trhr1* (*Trhr1*  $-/-$ ) prohibiting signaling of the pituitary gland to stimulate TSH production and therefore downregulate T3 and T4 production. Hyperthyroid mice were generated by three different approaches: 1) Stimulation with bovine TSH (500 mU) directly leads to the increased production of hormones in the thyroid. 2) Knock-out of *Trb1* (*Trb1*  $-/-$ ) inhibits the negative feedback loop and does not signal the hypothalamus and the pituitary to stop thyroid stimulation even though high thyroid hormone levels are present. 3) Knock-out of *Mct8* (*Mct8*  $-/-$ ) prevents thyroid hormones from leaving the thyrocytes into the bloodstream, thus inhibiting the negative feedback loop and leading to an overstimulation of the thyroid (see Materials and Methods section 2.2.1).

First, we asked whether *Tg* expression is increased in hypothyroid mice compared to control mice in response to hormone deficiency signaling. Second, we asked whether, on the contrary, *Tg* expression is decreased in thyrocytes of hyper-thyroid mice due to sufficient hormone levels in the blood and subsequent downregulation of hormone production. In both hypo- (Figure 3.23 A) and hyperthyroid (Figure 3.23 B-D) mice, however, we did not detect significant differences in the transcription level of *Tg* by qPCR. Consistent with this finding, there were no microscopically observable differences in the appearance of the TLs found in glands of hypo- and hyperthyroid mice compared to control mice (Figure 3.23 A-D). We therefore conclude that when the transcription of *Tg* and the resulting TL formation are initiated during thyrocyte differentiation, they are robustly maintained and not affected by the tested perturbations of thyroid hormone level.

Previous studies suggested a direct link between the presence of TSH and *Tg* gene activity in a cAMP dependent manner (Van Heuverswyn et al., 1984; Van Sande et al., 1990). In animals with thyroid glands overstimulated by TSH, no changes in the consistently high *Tg* transcription levels were found (Van Heuverswyn et al., 1984). In line with that, we did neither observe a change in the transcription level nor the TL appearance of the *Tg* gene in thyroid glands from hyperthyroid mice stimulated with TSH or deletions of *Trb1* or *Mct8*.

However, the study also showed that in animals with suppressed endogenous TSH levels, *Tg* transcription levels decreased (Van Heuverswyn et al., 1984). In contrast, we did not observe a variation in transcription level or TL appearance of the *Tg* gene in thyroid glands from hypothyroid mice with *Trhr1* deletion, even though this deletion ultimately also reduces TSH levels. These data indicate that *Trhr1*, in contrast to TSH itself as shown by other studies, does not regulate *Tg* expression directly. As suggested previously, *Tg* transcription control therefore might mostly rely directly on TSH.



**FIGURE 3.23: Perturbation of blood thyroid hormone levels does not affect *Tg* expression or TL formation.** Thyroid glands of hypo- (**A**) and hyperthyroid (**B-D**) mice were subjected to qPCR to determine the *Tg* expression level and to RNA FISH to assess TL formation (left and right, respectively). None of the perturbations in the thyroid hormone regulatory pathway affected the *Tg* transcription level or TL formation. (**A**) Hypothyroid mice were generated by deletion of thyrotropin releasing hormone receptor 1 (*Trhr1* -/-). The *Tg* transcription level as well as the appearance of the TL in thyrocytes from these mice is not different from control thyrocytes (*Trhr1* +/-). Hyperthyroid mice were generated by exogenous stimulation with (**B**) 500 mU bovine thyroid stimulating hormone (+ TSH), or by (**C**) deletion of both alleles of either thyroid hormone receptor beta (*Trb1* -/-) or (**D**) mono-carboxylate transporter 8 (*Mct8* -/-). The *Tg* transcription level as well as the appearance of the TL in thyrocytes from these mice is not different from control thyrocytes (- TSH, *Trb1* +/- and *Mct8* +/-). Scale bars 5  $\mu$ m.



### 3.7.5 MATURE *Tg* TRANSCRIPTS DO NOT RETAIN INTRONS

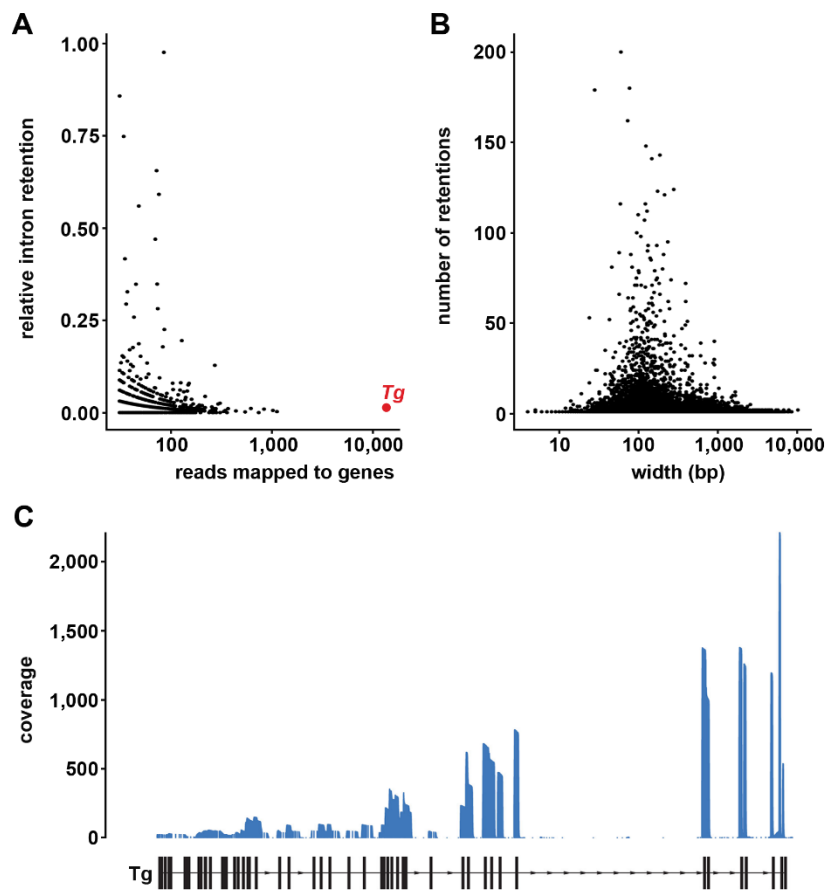
Results of this study point towards *Tg* as a gene that is constantly active. We assumed, however, that some level of transcriptional and/or translational fine-tuning might be present in thyrocytes to ensure appropriate TG levels.

First, we asked whether *Tg* transcription follows a circadian rhythm. A recent study in baboons indicates that *Tg* expression does not rely on a circadian rhythm (Mure et al., 2018). Out of  $\approx 11,000$  universally expressed genes in baboons almost 97 % exhibit rhythmic transcripts in at least one tissue (82 % of protein coding genes) (Mure et al., 2018). Globally, there seems to be a trend for most genes reaching their peak expression level during early morning or late afternoon. Using the dataset from this study, we checked the expression level of *Tg* in the thyroid at 12 different timepoints throughout the day and found stable *Tg* expression during these 24 h that does not seem to depend on a circadian rhythm.

Since we did not find evidence for the regulation of TG production at the level of gene activity, we reasoned that regulation might occur on the level of translation. Recent studies have pointed towards a phenomenon called intron retention as a mechanism to regulate mRNA levels in the cytoplasm (Braunschweig et al., 2014; Mauger et al., 2016; Naro and Sette, 2017; Wong et al., 2016). For example, the retention of introns in an mRNA after its dissociation from RNA Pol II prevents its export from the nucleus and thereby delays its translation. Only upon certain stimuli, rapid splicing is induced and mature mRNAs in turn undergo nuclear export (Mauger et al., 2016). Intron retention is therefore considered a “transcript tuner” and well documented for plants, fungi and insects, and also for mammals (Braunschweig et al., 2014; Kim et al., 2007; McGuire et al., 2008; Monteuuis et al., 2019; Naro and Sette, 2017). The nucleoplasmic RNA granules found within thyrocyte nuclei (see Results section 3.2.4, Figure 3.6 A) might represent such unspliced mRNAs remaining in the nucleus for regulation purposes. To examine whether unspliced introns are present in mature mRNAs in *Tg*, we performed direct Nanopore sequencing of RNA isolated from mouse thyrocytes in collaboration with LAFUGA (Gene Center, LMU München, Munich, Germany) and Dr. Sebastian Bultmann (BioSysM, LMU München, Munich, Germany). This recently developed technique has two major advantages. First, it enables sequencing of long reads which is especially important in case of the *Tg* gene with a length of 180 kb. Second, it does not rely on reverse transcription of RNA and therefore reduces bias introduced by cDNA amplification as it can occur during the preparation of libraries for Illumina sequencing (Aird et al., 2011).

Analysis of Nanopore RNA sequencing of RNA extracted from mouse thyroid showed that intron retention is more common for (i) weakly expressed genes (Figure 3.24 A) and (ii) for short introns (Figure 3.24 B).

Consistent with that, we did not find intron retention in the *Tg* mRNA (Figure 3.24 C). Even though the quality of the sequencing declines towards the TSS (as indicated by fewer reads in the first half of *Tg*, Figure 3.24 C) as a result of sequencing from the 3' end of isolated RNAs, this data strongly indicate that fine tuning of TG production is likely not regulated by intron retention in *Tg* mRNAs.



**FIGURE 3.24: Mouse *Tg* does not exhibit intron retention.** Direct Nanopore sequencing of total thyroid RNA shows no intron retention in the *Tg* gene, but reveals that intron retention is more common for lowly expressed genes and short introns. **(A)** Genes with few mapped reads (less abundant transcripts) show a higher level of relative intron retention. *Relative intron retention* was calculated as  $(\text{reads mapped to introns}) / (\text{reads mapped to genes})$ . **(B)** Shorter introns are more often retained than long introns. *Intron width* is depicted on the x-axis, the total number of retentions of the respective intron is depicted on the y-axis. **(C)** No intron retention is observed for *Tg*. Coverage decreases towards the TSS due to sequencing from the 3' end.



---

## 4 DISCUSSION AND FUTURE PERSPECTIVES

Despite the fact that transcription is a crucial step in gene expression and therefore extensively studied, the spatial organization of transcription is still poorly understood. On the one hand, numerous studies have elucidated players that fine-tune transcriptional regulation on the molecular scale – among them promoter-enhancer loops (Schoenfelder and Fraser, 2019), modifications of the RNA Pol II CTD (Eick and Geyer, 2013) and the complex dynamic network of factors associating with Pol II during different steps of transcription progression (Chen et al., 2018a) as well as RNA maturation (Muller-McNicoll and Neugebauer, 2013). On the other hand, studies on a broader scale showed that active genes reside in the nuclear interior and tend to locate at the edges of their harboring CT (Shah et al., 2018; Takizawa et al., 2008; Volpi et al., 2000). However, knowledge on the intermediate level of transcriptional organization, i.e. the three dimensional structure of a single transcribed gene, is surprisingly limited.

Our study contributes to filling this gap in knowledge by investigating the spatial organization of long genes. Thus, we circumvented the resolution limit of conventional light microscopy, which has prevented resolving the structure of small genes in the past (Lawrence et al., 1988; Trask et al., 1989).

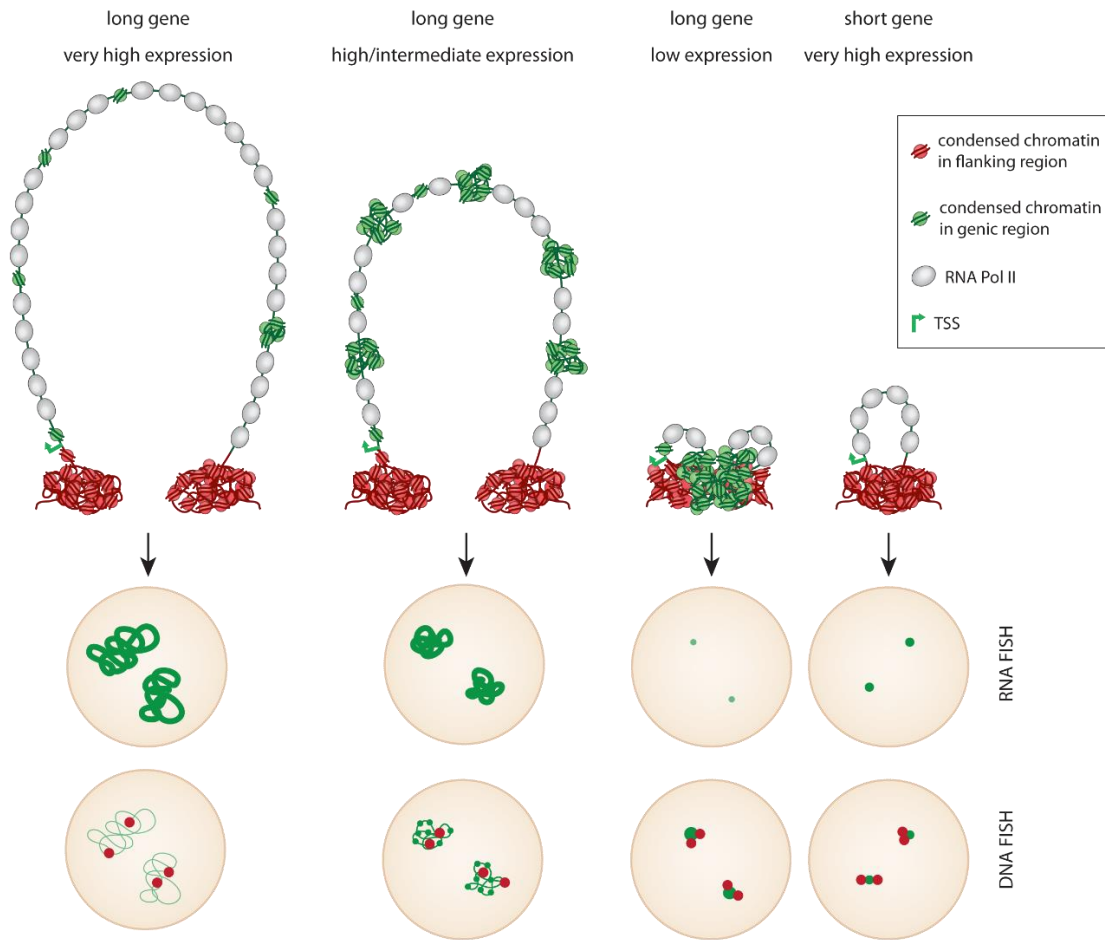
We showed that long and highly expressed genes are strongly decondensed and form microscopically detectable loop-like structures strongly resembling the lateral loops formed by highly expressed transcription units of LBCs (Figure 3.4). We termed these loops “transcription loops” (TLs) and determined two conditions crucial for their formation: genes need to be long and highly transcribed (Figures 3.3 and 3.4). Further, we showed that TLs represent sites of ongoing transcription as we can follow transcription along the loops using FISH probes specifically labeling exons (Figure 3.5). Using probes highlighting introns, we demonstrated that TLs exhibit co-transcriptional splicing (Figure 3.6 A,C). Additionally, we showed that TLs overlap with or are localized adjacent to nuclear speckles, in agreement with a high rate of ongoing splicing activity (Figure 3.6 B). We revealed that genes forming TL expand from their harboring CTs and dynamically modify surrounding chromosomal loci by separating flanking regions (Figures 3.12 and 3.13). We put forward the hypothesis that TL expansion is caused by an increased stiffness of the gene axis due to a dense decoration with nRNPs and provided evidence supporting this hypothesis.

#### 4.1 TRANSCRIPTIONAL BURSTING SHAPES TRANSCRIPTION LOOP STRUCTURE

Transcription loops exhibit different appearances (in shape and size) depending on the gene and cell type (Figures 3.4 and 3.10). The major determinant of TL appearance seems to be a gene's transcription rate. TLs formed by *Tg* in thyrocytes are the largest among the investigated genes. The notably longer *Ttn* and *Neb* genes, however, form smaller TLs compared to *Tg* TLs. This can be explained by the fact that expression levels of both muscle genes are  $\approx 10$ - to 20-fold lower than that of *Tg*. Similar observations were made on smooth muscle specific genes: although the *Cald1* gene is almost twice as long as the *Myh11* gene, it is expressed lower and thus forms smaller TLs. Based on these considerations, we conclude that there is an inverse correlation of transcription level and gene body condensation: the higher the transcription level, the higher the degree of gene decondensation, and hence the larger the formed TL.

We reason that the differences in gene condensation patterns reflect different bursting kinetics of the five genes. Transcription bursts are characterized by their initiation rate (frequency of transcription initiation) and duration (reflected by number of polymerases in one burst) (Tantale et al., 2016). The distance between the polymerases in one burst also varies depending on the gene (Tantale et al., 2016). We therefore suppose that the bursting kinetics of a gene's promoter strongly influences the appearance of the respective transcription loop: highly expressed genes are expressed in long bursts interrupted by short pauses resulting in a high degree of gene body decondensation; lowly expressed genes are transcribed in shorter bursts separated by long pauses resulting in a lower level of gene body decondensation (Figure 4.1).

It has to be noted, that a comparison of expression levels of different genes in different cell types is challenging due to high sample heterogeneity (Blake et al., 2020). It is therefore basically impossible to make statements about absolute differences in transcription levels between different genes in different tissues. However, in our study, for each cell type, four independent replicates were used for RNA-seq and the expression level of genes was calculated in transcripts per million (TPM) to ensure the highest possible comparability between samples. Thus, the data most probably still reflect relative expression levels, i.e. very highly expressed genes will yield very high TPM values, genes with an intermediate expression level will yield intermediate values and so forth, allowing a general comparison between expression level and TL structure. Using spike-in controls in RNA-seq experiments could circumvent these challenges (Chen et al., 2015).



**FIGURE 4.1: Transcription loop structure depends on a gene's bursting kinetics.** Top: Scheme depicting the proposed different bursting kinetics of very highly expressed long genes (e.g. *Tg* in thyrocytes), highly to moderately expressed long genes (e.g. *Ttn* and *Neb* in myotubes and *Myh11* and *Cald1* in smooth muscle), lowly expressed long genes (e.g. *Dmd* in myotubes) and highly expressed short genes (e.g. *Acta1* in myotubes) (left to right) reflecting the size of the respective TL. Very highly expressed long genes are transcribed in long frequent bursts resulting in the gene body almost entirely being covered by actively elongating polymerases (gray) with only very few condensed regions in between (green). Genes with a comparable length but slightly lower expression levels form smaller TLs since their gene body is covered by fewer and shorter bursts and they exhibit more condensed regions throughout the gene body. Lowly expressed long genes are covered by few polymerases and therefore do not show resolvable transcription loops. Likewise, short but highly expressed genes do not form resolvable TLs due to their size, even if the gene body is entirely covered with elongating polymerases. Bottom: Expected nuclear signals yielded by FISH for the respective genes. Long and very highly expressed genes form huge transcription loops detected by RNA FISH (green). In DNA FISH, the gene body (green) is hardly traceable due to the high level of decondensation. The gene flanking regions (red) are mostly located at a distance. Long genes with a high/intermediate expression level still exhibit TLs in RNA FISH, albeit their size is slightly smaller. In DNA FISH, these genes show a 'beads-on-a-string' pattern with alternating condensed and decondensed regions. Also in this case, gene flanking regions are mostly located at a distance. Long lowly expressed genes yield only a very faint RNA FISH signal and a spot-like DNA FISH signal due to the high degree of gene body condensation. In this case, gene flanks are in close proximity. For very highly expressed short genes, both RNA and DNA FISH yield a strong dot-like signal and gene flanks are in close proximity.

As shown previously, many factors such as DNA looping, transcription factor availability as well as general local chromatin environment regulate transcriptional bursting (Bartman et al., 2016; Kafri et al., 2016; Kalo et al., 2015; Larson et al., 2013; Nicolas et al., 2017). Many of these molecular features even specifically affect certain aspects of bursting (Nicolas et al., 2017). For example, DNA looping and nucleosome occupancy

modulate burst frequency, the number as well as the affinity of cis-regulatory elements modulate burst size, and TF availability and histone modifications affect both burst frequency and size (Nicolas et al., 2017). In *Drosophila*, changes in enhancer position and sequence specifically influence burst frequency (Fukaya et al., 2016). Mapping of enhancer regions of TL forming genes (especially *Tg* and *Tm*) via 4C or promoter capture Hi-C will further shed light on how the transcription of these genes is regulated. Additionally, the influence of promoter-enhancer loop kinetics can be assessed by single cell 4C as it allows to determine whether promoter-enhancer contacts are stable or transient. Contact frequencies in turn could then be correlated to transcription level and TL structure.

Visualization of TLs in living cells would provide a direct approach allowing to study transcriptional bursting. The viral derived MS2/MCP or PP7/PCP systems, relying on the incorporation of stem loop structures into RNA and their detection via the respective fluorescently labeled binder (Bertrand et al., 1998; Larson et al., 2011; Lim et al., 2001; Peabody, 1993), may provide promising tools to visualize nascent transcripts and investigate transcriptional bursting and TL formation in living cells. To this end, the stem loop repeats would need to be incorporated into the 5' untranslated region (UTR) of the respective gene in order to visualize RNAs along the entire transcription loop, since 3' UTR incorporation would label nascent RNAs only at the end of the transcript. Specifically targeting an intron with MS2/MCP or PP7/PCP potentially enables visualizing co-transcriptional splicing.

Possibly, TLs may as well be visualized by labeling active RNA Pol II due to its high abundance. Previous studies involving live cell imaging of RNA Pol II mostly relied on expressing an exogenous  $\alpha$ -amanitin resistant fluorescent fusion form of Pol II and simultaneous degradation of endogenous Pol II by  $\alpha$ -amanitin (Cho et al., 2016a; Hongo et al., 2008; Nguyen et al., 1996). Recently, a CRISPR/Cas9 based gene editing approach was successfully used to endogenously incorporate a fluorescent protein sequence into the *Rpb1* locus (Cho et al., 2016b; Steurer et al., 2018). Another promising approach is labeling RNA Pol II with a photoswitchable fluorescent protein as described previously by Cho and colleagues (Cho et al., 2016b), since this does not only permit super resolution imaging but could provide further insights into Pol II dynamics on TLs. Converting a small fraction of fluorescent protein fused to initiating Pol II at the TSS would allow to follow polymerase elongation along the gene body and thus may provide conclusions on Pol II speed as well as a gene's bursting kinetics.

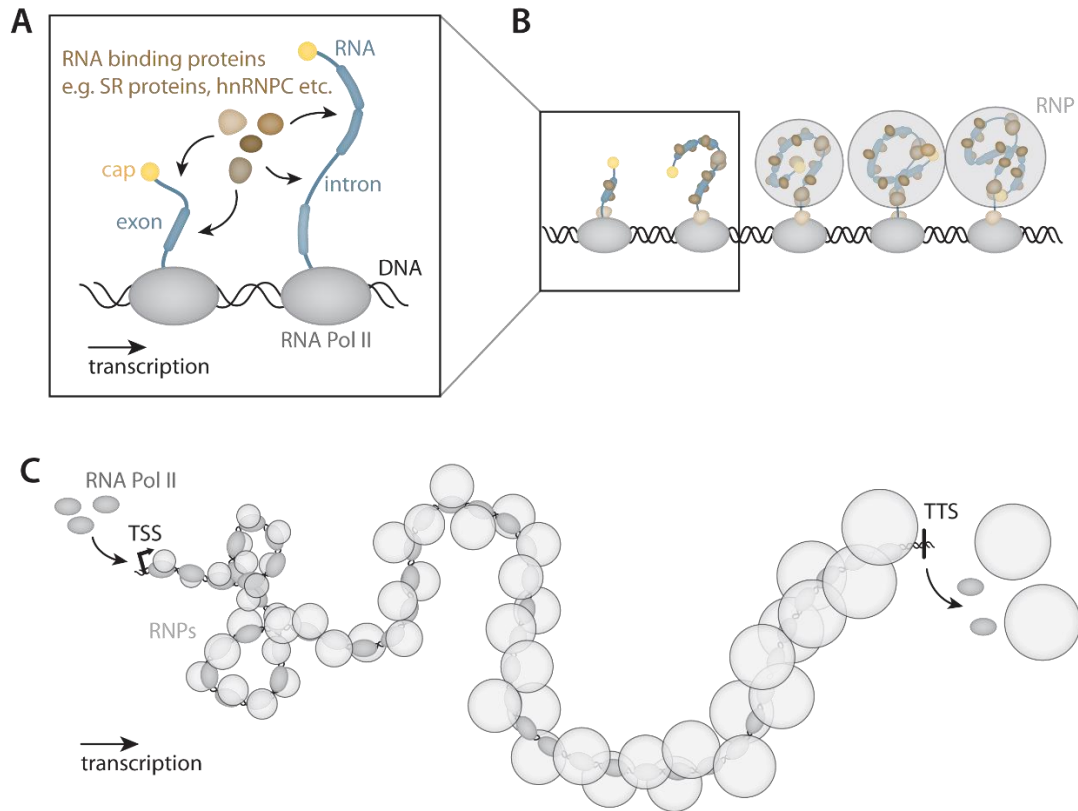
The mintbody strategy (Sato et al., 2013) could be used to visualize the *Tm* locus together with occupying polymerases in Pmi28 myotubes. This system relies on the detection of the gene via dCas9-sfGFP and the simultaneous detection of RNA Pol II via a modification-specific intracellular antibody (mintbody) specific for the elongating serine 2 phosphorylated (S2P) form of the Pol II CTD (Sato et al., 2013). Given that the specificity of *Tm* labeling via dCas9-GFP as well as the specificity of the mintbody for elongating RNA Pol II are assured, simultaneous labeling of nascent RNAs and elongating Pol II could further elucidate the spatio-temporal dynamics of TL formation.

## 4.2 TRANSCRIPTION LOOPS ARE STIFF STRUCTURES DUE TO DENSE DECORATION WITH RIBONUCLEOPROTEIN PARTICLES

Nascent transcripts are coated by a complex network of partially inter-dependent RNA binding proteins (RBPs) (Figure 4.2 A) and form compact nascent ribonucleoprotein particles (nRNPs) (Figure 4.2 B) (Muller-McNicoll and Neugebauer, 2013). RNA recognition motif (RRM) containing SR-proteins – among them slicing factor SC35 – associate with pre-mRNAs to promote (alternative) splicing (Shepard and Hertel, 2009). Exon junction complex (EJC) forms on sites of joined exons during splicing and acts as a binding platform for peripheral factors (Le Hir et al., 2016). hnRNPC binds introns in nascent transcripts and greatly influences pre-mRNA compaction (Konig et al., 2010). It has been proposed that hnRNPC tetramers serve as ‘RNA nucleosome’ as they harbor 150–250 nt of RNA wrapped around them, likewise to DNA wrapped around histones (Huang et al., 1994; McAfee et al., 1996). Transcription export complex (TREX), linking transcription and mRNA export, is already co-transcriptionally recruited to nascent transcripts in a splicing-dependent manner in mammals (Reed and Cheng, 2005). Thus, we hypothesized that this dense occupation with bulky RNP entities renders the gene body stiff (Figure 4.2 C).

We provide several lines of evidence indicating that TLs are stiff structures. First, we showed that gene flanks are separated in a transcription dependent manner and that genes loop out from their harboring chromosome (Figure 3.13). In agreement with this, lowly expressed genes, such as *Neb* and *Cald1*, do not loop far from their CTs. Direct evidence for the splitting of chromosome 15 by the *Tg* TL as observed by FISH, was not found by Hi-C analysis. We tentatively explain this by (i) the relatively high heterogeneity of the thyroid Hi-C sample — we estimate that about 40 % of cells in the tissue are non-thyocytes — and (ii) the appearance of CT splitting in only 2 % of homologs that might escape detection by Hi-C.

Second, *Tg* TLs exhibit different degrees of coiling on different parts of the gene. *Tg*’s long intron 41, measuring  $\approx 53$  kb, is generally more expanded than the rest of the gene. We assume that the decoration of the intron with nRNPs, formed by nRNAs growing from approx. 7 kb (at the beginning of intron 41) to approx. 61 kb (at the end of intron 41), particularly increases the stiffness of the TL and thus the radius of curvature in this region. Since RNP size correlates with the length of the RNA component (Beyer et al., 1977; Derman et al., 1976), RNPs close to the TSS are (much) smaller than RNPs close to the TTS. Gene bodies of transcribed genes will therefore presumably generally exhibit stronger coiling towards the 5’ end where the RNPs are small and enable a smaller bending radius, whereas they will exhibit strong expansion towards their 3’ ends where RNPs are large and prevent high curvatures (Figure 4.2 C). Especially in the case of *Ttn*, a gene with an unusually high exon/intron ratio and thus with an exceptionally long mRNA of  $\approx 103$  kb, the nRNPs increase in size towards the TTS, despite co-transcriptional splicing, which is indeed reflected by a decrease in the *cis/trans* contact frequency ratio towards the 3’ end of the *Ttn* gene (Figure 3.13). In line with our hypothesis, small genes such as *Acta1* did not show TL formation but only a dot-like FISH signal even though its expression level is very high. Based on our hypothesis, we reason that due to the small size of *Acta1* and therefore the small size of nascent RNPs, the gene is highly coiled which prevents its structural resolution.



**FIGURE 4.2: Possible mechanism of transcription loop formation.** (A) Nascent transcripts are bound by hundreds of RNA binding proteins (RBPs) for splicing, quality control, nuclear export, etc. (B) Nascent transcripts and associated RBPs are tightly packed and form so-called ribonucleoprotein particles (RNPs). (C) Densely spaced polymerases carry RNPs along the gene body of a highly transcribed gene. The size of the RNPs increases towards the end of the gene as the associated RNA grows progressively and determines the rigidity of the respective part of a TL. Close to the TSS, RNPs are small and therefore allow the gene body to coil. Close to the TTS, RNPs are large and bulky and render the gene body stiff due to their intrinsic steric configuration. In turn, this leads to the formation of TLs on long and highly expressed genes. *Splicing as well as transcriptional bursting are not depicted for simplicity reasons.* TSS: transcription start site; TTS: transcription termination site.

The granules seen in myotube nuclei by electron microscopy (EM) (Figure 3.17) likely correspond to nascent RNPs decorating the *Ttn* gene. They resemble RNPs previously visualized on Balbiani rings of polytene chromosomes (Bjork and Wieslander, 2015) and display a heterogeneity in size, probably reflecting their growth. Further studies of these granular structures are currently on the way in our group and will serve to (i) further corroborate the hypothesis that their formation depends on active transcription by imaging myotube nuclei treated with transcription inhibitors and (ii) to determine that granules indeed correspond to the *Ttn* gene body by combining EM and FISH in a correlative approach.

Since methods to measure DNA flexibility *in vivo* are highly artificial and only provide general estimations on DNA stiffness in very defined environments (Peters and Maher, 2010), we chose to rely on polymer models of chromatin to investigate physical parameters of TL formation. Indeed, polymer models strongly corroborated our hypothesis that gene body stiffness substantially contributes to TL formation (Figure 3.18).

### 4.3 TRANSCRIPTION LOOP FORMATION AS UNIVERSAL MODE OF TRANSCRIPTION

Our hypothesis suggests that TL formation is the universal principle of transcription in eukaryotes. In favor of this hypothesis, TLs have been observed in all vertebrate species (Figure 3.19) and in native tissues (Figures 3.4, 3.10 and 3.19) as well as cultured cells (Figures 3.4, 3.9 and 3.10). We assume that all eukaryotic genes form TLs when they are transcribed. We suppose that until now TLs have escaped visual detection by FISH because the investigated genes were either too small or too lowly expressed to yield a resolvable signal.

Notably, even though we show TL formation for five genes and in diverse cell types and different vertebrates, other eukaryotic species such as plants were not considered in this study. Therefore, to support our hypothesis, the investigation of more examples would be necessary. However, as our search for long and highly expressed genes in the published GTEx RNA-seq dataset resulted in very few matches, the major limitation in this concern will probably be the low number of suitable candidate genes.

Therefore, experimental approaches aiming at increasing either the length of highly expressed genes or the transcription level of long genes could be strategies to further accumulate evidence in favor of our hypothesis. Various other dCas9-based approaches were successfully applied previously to upregulate the transcription of certain genes and could be used to complement our experiments. For example, the SunTag activation system relies on a dCas9 fused to a repeating polypeptide scaffold (SuperNova tag, SunTag) which recruits antibodies fused to an activation domain (Tanenbaum et al., 2014). Since multiple antibodies bind to one scaffold, transcription activator recruitment is thus amplified. The synergistic activation mediator (SAM) system relies on recruitment of several transcriptional factors via hairpin aptamers that are incorporated into the gRNA yielding beneficial concerted action of different activators (Konermann et al., 2015; Zhang et al., 2015). Furthermore, a short highly expressed gene (such as *Acta1*) could be elongated by CRISPR/Cas-mediated insertion of DNA stretches in order to form resolvable TLs.

The finding that long and highly expressed genes form TLs urges the need for revisiting current models for the spatial organization of transcription. Our study indicates that transcription is carried out by RNA polymerases moving along template DNA like a train on a track accompanied by their attached RNP cargo. Our findings set the ground for reconsiderations on the basic principles of transcription and highlight the significance as well as the versatility of TL forming genes for investigating these principles.





---

## 5 APPENDIX

### 5.1 REFERENCES

- Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* *12*, R18.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., *et al.* (2014). An atlas of active enhancers across human cell types and tissues. *Nature* *507*, 455-461.
- Anton, T., Bultmann, S., Leonhardt, H., and Markaki, Y. (2014). Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* *5*, 163-172.
- Antonica, F., Kasprzyk, D.F., Opitz, R., Iacovino, M., Liao, X.H., Dumitrescu, A.M., Refetoff, S., Peremans, K., Manto, M., Kyba, M., *et al.* (2012). Generation of functional thyroid from embryonic stem cells. *Nature* *491*, 66-71.
- Bagnoli, J.W., Ziegenhain, C., Janjic, A., Wange, L.E., Vieth, B., Parekh, S., Geuder, J., Hellmann, I., and Enard, W. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCR-seq. *Nat Commun* *9*, 2937.
- Baldi, S., Korber, P., and Becker, P.B. (2020). Beads on a string-nucleosome array arrangements and folding of the chromatin fiber. *Nat Struct Mol Biol* *27*, 109-118.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res* *21*, 381-395.
- Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* *280*, 17732-17736.
- Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* *410*, 120-124.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* *129*, 823-837.
- Bartman, C.R., Hsu, S.C., Hsiung, C.C., Raj, A., and Blobel, G.A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell* *62*, 237-247.
- Belaghal, H., Dekker, J., and Gibcus, J.H. (2017). Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods* *123*, 56-65.
- Bernardi, G. (1989). The isochore organization of the human genome. *Annu Rev Genet* *23*, 637-661.
- Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. *Mol Cell* *2*, 437-445.
- Beyer, A.L., Christensen, M.E., Walker, B.W., and LeStourgeon, W.M. (1977). Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell* *11*, 127-138.

- Bienko, M., Crosetto, N., Teytelman, L., Klemm, S., Itzkovitz, S., and van Oudenaarden, A. (2013). A versatile genome-scale PCR-based pipeline for high-definition DNA FISH. *Nat Methods* *10*, 122-124.
- Birnbaum, R.Y., Clowney, E.J., Agamy, O., Kim, M.J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S.L., Wenger, A.M., Nguyen, L., *et al.* (2012). Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res* *22*, 1059-1068.
- Bjork, P., and Wieslander, L. (2015). The Balbiani Ring Story: Synthesis, Assembly, Processing, and Transport of Specific Messenger RNA-Protein Complexes. *Annu Rev Biochem* *84*, 65-92.
- Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., *et al.* (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat Struct Mol Biol* *25*, 833-840.
- Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., *et al.* (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* *175*, 1842-1855 e1816.
- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A., *et al.* (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* *171*, 557-572 e524.
- Boyle, S., Gilchrist, S., Bridger, J.M., Mahy, N.L., Ellis, J.A., and Bickmore, W.A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* *10*, 211-219.
- Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* *24*, 1774-1786.
- Brent, G.A. (2012). Mechanisms of thyroid hormone action. *J Clin Invest* *122*, 3035-3043.
- Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A., Smith, A.J., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R., *et al.* (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J Cell Biol* *182*, 1083-1097.
- Buckley, M.S., and Lis, J.T. (2014). Imaging RNA Polymerase II transcription sites in living cells. *Curr Opin Genet Dev* *25*, 126-130.
- Bushnell, D.A., Cramer, P., and Kornberg, R.D. (2002). Structural basis of transcription: alpha-amanitin-RNA polymerase II cocystal at 2.8 Å resolution. *Proc Natl Acad Sci U S A* *99*, 1218-1222.
- Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature* *440*, 358-362.
- Callan, H.G. (1986). Lampbrush chromosomes. *Mol Biol Biochem Biophys* *36*, 1-252.
- Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol Cell* *49*, 825-837.
- Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., E, P.R.I., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., *et al.* (2015). Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* *12*, 326-328.
- Chen, F.X., Smith, E.R., and Shilatifard, A. (2018a). Born to run: control of transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* *19*, 464-478.

- Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J., and Belmont, A.S. (2018b). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J Cell Biol* *217*, 4025-4048.
- Cho, W.K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.H., Lavis, L.D., Lionnet, T., *et al.* (2016a). RNA Polymerase II cluster dynamics predict mRNA output in living cells. *Elife* *5*.
- Cho, W.K., Jayanth, N., Mullen, S., Tan, T.H., Jung, Y.J., and Cisse, II (2016b). Super-resolution imaging of fluorescently labeled, endogenous RNA Polymerase II in living cells with CRISPR/Cas9-mediated gene editing. *Sci Rep* *6*, 35949.
- Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, II (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* *361*, 412-415.
- Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., *et al.* (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* *361*.
- Chu, M., Gregorio, C.C., and Pappas, C.T. (2016). Nebulin, a multi-functional giant. *J Exp Biol* *219*, 146-152.
- Chuang, C.H., Carpenter, A.E., Fuchsova, B., Johnson, T., de Lanerolle, P., and Belmont, A.S. (2006). Long-range directional movement of an interphase chromosome site. *Curr Biol* *16*, 825-831.
- Chubb, J.R., Trcek, T., Shenoy, S.M., and Singer, R.H. (2006). Transcriptional pulsing of a developmental gene. *Curr Biol* *16*, 1018-1025.
- Cisse, II, Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* *341*, 664-667.
- Cramer, P. (2019). Organization and regulation of gene transcription. *Nature* *573*, 45-54.
- Cremer, M., Grasser, F., Lanctot, C., Muller, S., Neusser, M., Zinner, R., Solovei, I., and Cremer, T. (2008). Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol Biol* *463*, 205-239.
- Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C., and Cremer, T. (2001). Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res* *9*, 541-567.
- Cremer, T., and Cremer, M. (2010). Chromosome territories. *Cold Spring Harb Perspect Biol* *2*, a003889.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* *107*, 21931-21936.
- Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* *145*, 1119-1131.
- Daneholt, B. (1975). Transcription in polytene chromosomes. *Cell* *4*, 1-9.
- de Wit, E., Vos, E.S., Holwerda, S.J., Valdes-Quezada, C., Verstegen, M.J., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* *60*, 676-684.

- Derman, E., Goldberg, S., and Darnell, J.E., Jr. (1976). hnRNA in HeLa cells: distribution of transcript sizes estimated from nascent molecule profile. *Cell* *9*, 465-472.
- Deutsch, M., and Long, M. (1999). Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res* *27*, 3219-3228.
- Di Jeso, B., and Arvan, P. (2016). Thyroglobulin From Molecular and Cellular Biology to Clinical Endocrinology. *Endocr Rev* *37*, 2-36.
- Diez, J.L. (1973). Effect of cordycepin (3'-desoxyadenosine) on polytene chromosomes of *Chironomus pallidivittatus* salivary glands. *Chromosoma* *42*, 345-358.
- DiStefano, J.J. (1973). The thyroid hormone feedback control system. *Engineering Principles in Physiology*, 261 - 289.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* *485*, 376-380.
- Dossin, F., Pinheiro, I., Zylitz, J.J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y., *et al.* (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*.
- Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y., Beauchamp, K.A., Wang, L.P., Simmonett, A.C., Harrigan, M.P., Stern, C.D., *et al.* (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* *13*, e1005659.
- Eckersley-Maslin, M.A., and Spector, D.L. (2014). Random monoallelic expression: regulating gene expression one allele at a time. *Trends Genet* *30*, 237-244.
- Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* *113*, 8456-8490.
- Eskiw, C.H., and Fraser, P. (2011). Ultrastructural study of transcription factories in mouse erythroblasts. *J Cell Sci* *124*, 3676-3683.
- Eskiw, C.H., Rapp, A., Carter, D.R., and Cook, P.R. (2008). RNA polymerase II activity is located on the surface of protein-rich transcription factories. *J Cell Sci* *121*, 1999-2007.
- Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B.R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I., *et al.* (2019). Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* *570*, 395-399.
- Flavahan, W.A., Drier, Y., Liao, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* *529*, 110-114.
- Franke, W.W., Scheer, U., Trendelenburg, M.F., Spring, H., and Zentgraf, H. (1976). Absence of nucleosomes in transcriptionally active chromatin. *Cytobiologie* *13*, 401-434.
- Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell* *166*, 358-368.
- Gendrel, A.V., Attia, M., Chen, C.J., Diabangouaya, P., Servant, N., Barillot, E., and Heard, E. (2014). Developmental dynamics and disease potential of random monoallelic gene expression. *Dev Cell* *28*, 366-380.
- Georgomanolis, T., Sofiadis, K., and Papantonis, A. (2016). Cutting a Long Intron Short: Recursive Splicing and Its Implications. *Front Physiol* *7*, 598.

- Ghamari, A., van de Corput, M.P., Thongjuea, S., van Cappellen, W.A., van Ijcken, W., van Haren, J., Soler, E., Eick, D., Lenhard, B., and Grosveld, F.G. (2013). In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes Dev* 27, 767-777.
- Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. *Mol Cell* 49, 773-782.
- Goetze, S., Mateos-Langerak, J., Gierman, H.J., de Leeuw, W., Giromus, O., Indemans, M.H., Koster, J., Ondrej, V., Versteeg, R., and van Driel, R. (2007). The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol* 27, 4475-4487.
- Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* 123, 1025-1036.
- Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. *Nat Rev Genet* 8, 35-46.
- Groba, C., Mayerl, S., van Mullem, A.A., Visser, T.J., Darras, V.M., Habenicht, A.J., and Heuer, H. (2013). Hypothyroidism compromises hypothalamic leptin signaling in mice. *Mol Endocrinol* 27, 586-597.
- Grunwald, D., and Singer, R.H. (2010). In vivo imaging of labelled endogenous beta-actin mRNA during nucleocytoplasmic transport. *Nature* 467, 604-607.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., *et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948-951.
- Guenatri, M., Bailly, D., Maison, C., and Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol* 166, 493-505.
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., *et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162, 900-910.
- Habermann, F.A., Cremer, M., Walter, J., Kreth, G., von Hase, J., Bauer, K., Wienberg, J., Cremer, C., Cremer, T., and Solovei, I. (2001). Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res* 9, 569-584.
- Hampsey, M., Singh, B.N., Ansari, A., Laine, J.P., and Krishnamurthy, S. (2011). Control of eukaryotic gene expression: gene loops and transcriptional memory. *Adv Enzyme Regul* 51, 118-125.
- Hebenstreit, D. (2013). Are gene loops the cause of transcriptional noise? *Trends Genet* 29, 333-338.
- Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* 18, 637-650.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934-947.
- Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* 169, 13-23.
- Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., *et al.* (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 351, 1454-1458.
- Hoermann, R., Midgley, J.E., Larisch, R., and Dietrich, J.W. (2015). Homeostatic Control of the Thyroid-Pituitary Axis: Perspectives for Diagnosis and Treatment. *Front Endocrinol (Lausanne)* 6, 177.

- Hongo, E., Ishihara, Y., Sugaya, K., and Sugaya, K. (2008). Characterization of cells expressing RNA polymerase II tagged with green fluorescent protein: effect of ionizing irradiation on RNA synthesis. *Int J Radiat Biol* *84*, 778-787.
- Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol Cell* *48*, 471-484.
- Hu, Y., Kireev, I., Plutz, M., Ashourian, N., and Belmont, A.S. (2009). Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. *J Cell Biol* *185*, 87-100.
- Huang, M., Rech, J.E., Northington, S.J., Flicker, P.F., Mayeda, A., Krainer, A.R., and LeStourgeon, W.M. (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.
- Huber, P.A. (1997). Caldesmon. *Int J Biochem Cell Biol* *29*, 1047-1051.
- Hyman, A.A., Weber, C.A., and Julicher, F. (2014). Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* *30*, 39-58.
- Iborra, F.J., Pombo, A., Jackson, D.A., and Cook, P.R. (1996). Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J Cell Sci* *109* (Pt 6), 1427-1436.
- Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* *9*, 999-1003.
- Iwen, K.A., Schroder, E., and Brabant, G. (2013). Thyroid hormones and the metabolic syndrome. *Eur Thyroid J* *2*, 83-92.
- Jackson, D.A., Hassan, A.B., Errington, R.J., and Cook, P.R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J* *12*, 1059-1065.
- Jackson, D.A., Iborra, F.J., Manders, E.M., and Cook, P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* *9*, 1523-1536.
- Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* *293*, 1074-1080.
- Johansson, H.E., Liljas, L., and Uhlenbeck, O.C. (1997). RNA recognition by the MS2 phage coat protein. *Sem Virol* *8*, 176-185.
- Kafri, P., Hasenson, S.E., Kanter, I., Sheinberger, J., Kinor, N., Yunger, S., and Shav-Tal, Y. (2016). Quantifying beta-catenin subcellular dynamics and cyclin D1 mRNA transcription during Wnt signaling in single living cells. *Elife* *5*.
- Kalo, A., Kanter, I., Shraga, A., Sheinberger, J., Tzemach, H., Kinor, N., Singer, R.H., Lionnet, T., and Shav-Tal, Y. (2015). Cellular Levels of Signaling Factors Are Sensed by beta-actin Alleles to Modulate Transcriptional Pulse Intensity. *Cell Rep* *11*, 419-432.
- Khanna, N., Hu, Y., and Belmont, A.S. (2014). HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr Biol* *24*, 1138-1144.
- Kim, E., Magen, A., and Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res* *35*, 125-131.

- Koehler, D., Zakhartchenko, V., Froenicke, L., Stone, G., Stanyon, R., Wolf, E., Cremer, T., and Brero, A. (2009). Changes of higher order chromatin arrangements during major genome activation in bovine preimplantation embryos. *Exp Cell Res* *315*, 2053-2063.
- Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* *517*, 583-588.
- Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* *17*, 909-915.
- Korenberg, J.R., and Rykowski, M.C. (1988). Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* *53*, 391-400.
- Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* *184*, 868-871.
- Kosak, S.T., Scalzo, D., Alworth, S.V., Li, F., Palmer, S., Enver, T., Lee, J.S., and Groudine, M. (2007). Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol* *5*, e309.
- Kulaeva, O.I., Hsieh, F.K., and Studitsky, V.M. (2010). RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. *Proc Natl Acad Sci U S A* *107*, 11325-11330.
- Kupper, K., Kolbl, A., Biener, D., Dittrich, S., von Hase, J., Thormeyer, T., Fiegler, H., Carter, N.P., Speicher, M.R., Cremer, T., *et al.* (2007). Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma* *116*, 285-306.
- Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* *155*, 1049-1060.
- Lancot, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet* *8*, 104-115.
- Larson, D.R., Frittsch, C., Sun, L., Meng, X., Lawrence, D.S., and Singer, R.H. (2013). Direct observation of frequency modulated transcription in single cells using light activation. *Elife* *2*, e00750.
- Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* *332*, 475-478.
- Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, A., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. *Nature* *565*, 251-254.
- Lawrence, J.B., Villave, C.A., and Singer, R.H. (1988). Sensitive, high-resolution chromatin and chromosome mapping in situ: presence and orientation of two closely integrated copies of EBV in a lymphoma line. *Cell* *52*, 51-61.
- Le Hir, H., Sauliere, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. *Nat Rev Mol Cell Biol* *17*, 41-54.
- Liang, G., Lin, J.C., Wei, V., Yoo, C., Cheng, J.C., Nguyen, C.T., Weisenberger, D.J., Egger, G., Takai, D., Gonzales, F.A., *et al.* (2004). Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc Natl Acad Sci U S A* *101*, 7357-7362.

- Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D.C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet* *80*, 224-234.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* *326*, 289-293.
- Lim, F., Downey, T.P., and Peabody, D.S. (2001). Translational repression and specific RNA binding by the coat protein of the Pseudomonas phage PP7. *J Biol Chem* *276*, 22507-22513.
- Link, S., Spitzer, R.M.M., Sana, M., Torrado, M., Volker-Albert, M.C., Keilhauer, E.C., Burgold, T., Punzeler, S., Low, J.K.K., Lindstrom, I., *et al.* (2018). PWWP2A binds distinct chromatin moieties and interacts with an MTA1-specific core NuRD complex. *Nat Commun* *9*, 4300.
- Linke, W.A. (2018). Titin Gene and Protein Functions in Passive and Active Muscle. *Annu Rev Physiol* *80*, 389-411.
- Little, J.W. (2006a). Thyroid disorders. Part I: hyperthyroidism. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* *101*, 276-284.
- Little, J.W. (2006b). Thyroid disorders. Part II: hypothyroidism and thyroiditis. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* *102*, 148-153.
- Lu, H., Yu, D., Hansen, A.S., Ganguly, S., Liu, R., Heckert, A., Darzacq, X., and Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* *558*, 318-323.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *389*, 251-260.
- Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., *et al.* (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* (London: Chapman and Hall).
- Macgregor, H.C. (2012). Chromomeres revisited. *Chromosome Res* *20*, 911-924.
- Maison, C., and Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* *5*, 296-304.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* *339*, 823-826.
- Martin, S., and Pombo, A. (2003). Transcription factories: quantitative studies of nanostructures in the mammalian nucleus. *Chromosome Res* *11*, 461-470.
- 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.
- Mayer, R., Brero, A., von Hase, J., Schroeder, T., Cremer, T., and Dietzel, S. (2005). Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol* *6*, 44.
- McAfee, J.G., Shahied-Milam, L., Soltaninassab, S.R., and LeSturgeon, W.M. (1996). A major determinant of hnRNP C protein binding to RNA is a novel bZIP-like RNA binding domain. *RNA* *2*, 1139-1152.



- McGuire, A.M., Pearson, M.D., Neafsey, D.E., and Galagan, J.E. (2008). Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol* **9**, R50.
- Mehlin, H., and Daneholt, B. (1993). The Balbiani ring particle: a model for the assembly and export of RNPs from the nucleus? *Trends Cell Biol* **3**, 443-447.
- Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., *et al.* (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **47**, 598-606.
- Miller, O.L., Jr., and Beatty, B.R. (1969a). Extrachromosomal nucleolar genes in amphibian oocytes. *Genetics* **61**, Suppl:133-143.
- Miller, O.L., Jr., and Beatty, B.R. (1969b). Visualization of nucleolar genes. *Science* **164**, 955-957.
- Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. *Cell* **128**, 787-800.
- Misteli, T., and Spector, D.L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell* **3**, 697-705.
- Monteuuis, G., Wong, J.J.L., Bailey, C.G., Schmitz, U., and Rasko, J.E.J. (2019). The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res* **47**, 11497-11513.
- Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A., and Berezney, R. (1996). A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci U S A* **93**, 8253-8257.
- Muller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat Rev Genet* **14**, 275-287.
- Mure, L.S., Le, H.D., Benegiamo, G., Chang, M.W., Rios, L., Jillani, N., Ngotho, M., Kariuki, T., Dkhissi-Benyahya, O., Cooper, H.M., *et al.* (2018). Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science* **359**.
- Naro, C., and Sette, C. (2017). Timely-regulated intron retention as device to fine-tune protein expression. *Cell Cycle* **16**, 1321-1322.
- Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta* **1809**, 34-45.
- Neusser, M., Schubel, V., Koch, A., Cremer, T., and Muller, S. (2007). Evolutionarily conserved, cell type and species-specific higher order chromatin arrangements in interphase nuclei of primates. *Chromosoma* **116**, 307-320.
- Nguyen, V.T., Giannoni, F., Dubois, M.F., Seo, S.J., Vigneron, M., Kedinger, C., and Bensaude, O. (1996). In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. *Nucleic Acids Res* **24**, 2924-2929.
- Nicolas, D., Phillips, N.E., and Naef, F. (2017). What shapes eukaryotic transcriptional bursting? *Mol Biosyst* **13**, 1280-1290.
- Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., and Thanos, D. (2009). Transcription factors mediate long-range enhancer-promoter interactions. *Proc Natl Acad Sci U S A* **106**, 20222-20227.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., *et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385.

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

Oetting, A., and Yen, P.M. (2007). New insights into thyroid hormone action. *Best Pract Res Clin Endocrinol Metab* **21**, 193-208.

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

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

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

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.

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.

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

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

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

Peabody, D.S. (1993). The RNA binding site of bacteriophage MS2 coat protein. *EMBO J* **12**, 595-600.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., *et al.* (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* **12**, 1577-1589.

Peters, J.P., 3rd, and Maher, L.J. (2010). DNA curvature and flexibility in vitro and in vivo. *Q Rev Biophys* **43**, 23-63.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., *et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281-1295.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., and Gray, J. (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl Acad Sci U S A* **85**, 9138-9142.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., *et al.* (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402-405.

Pott, S., and Lieb, J.D. (2015). What are super-enhancers? *Nat Genet* **47**, 8-12.

- Punzeler, S., Link, S., Wagner, G., Keilhauer, E.C., Kronbeck, N., Spitzer, R.M., Leidescher, S., Markaki, Y., Mentele, E., Regnard, C., *et al.* (2017). Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation. *EMBO J* **36**, 2263-2279.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283.
- Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309.
- Raj, A., Rifkin, S.A., Andersen, E., and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913-918.
- Rau, A., Gallopin, M., Celeux, G., and Jaffrezic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* **29**, 2146-2152.
- Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. *Curr Opin Cell Biol* **17**, 269-273.
- Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-432.
- Rieder, D., Trajanoski, Z., and McNally, J.G. (2012). Transcription factories. *Front Genet* **3**, 221.
- Ron, G., Globerson, Y., Moran, D., and Kaplan, T. (2017). Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nat Commun* **8**, 2237.
- Rudd, M.D., and Luse, D.S. (1996). Amanitin greatly reduces the rate of transcription by RNA polymerase II ternary complexes but fails to inhibit some transcript cleavage modes. *J Biol Chem* **271**, 21549-21558.
- Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., *et al.* (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**.
- Sakharkar, M.K., Chow, V.T., and Kanguene, P. (2004). Distributions of exons and introns in the human genome. *In Silico Biol* **4**, 387-393.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113.
- Sato, Y., Mukai, M., Ueda, J., Muraki, M., Stasevich, T.J., Horikoshi, N., Kujirai, T., Kita, H., Kimura, T., Hira, S., *et al.* (2013). Genetically encoded system to track histone modification in vivo. *Sci Rep* **3**, 2436.
- 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., Heintzmann, R., and Leonhardt, H. (2010). A guide to super-resolution fluorescence microscopy. *J Cell Biol* **190**, 165-175.
- Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **20**, 437-455.
- Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., *et al.* (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* **25**, 582-597.

- Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* *18*, 1251-1262.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* *148*, 458-472.
- Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.L., Koulana, N., Cronin, C., Karp, C., Liaw, E.J., *et al.* (2018). Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* *174*, 363-376 e316.
- Shandilya, J., and Roberts, S.G. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta* *1819*, 391-400.
- Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., *et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* *488*, 116-120.
- Shepard, P.J., and Hertel, K.J. (2009). The SR protein family. *Genome Biol* *10*, 242.
- Shibusawa, N., Hashimoto, K., Nikrodhanond, A.A., Liberman, M.C., Applebury, M.L., Liao, X.H., Robbins, J.T., Refetoff, S., Cohen, R.N., and Wondisford, F.E. (2003). Thyroid hormone action in the absence of thyroid hormone receptor DNA-binding in vivo. *J Clin Invest* *112*, 588-597.
- Shopland, L.S., Johnson, C.V., Byron, M., McNeil, J., and Lawrence, J.B. (2003). Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J Cell Biol* *162*, 981-990.
- Sibley, C.R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., Trabzuni, D., Ryten, M., Weale, M.E., Hardy, J., *et al.* (2015). Recursive splicing in long vertebrate genes. *Nature* *521*, 371-375.
- Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* *16*, 1128-1133.
- Sobell, H.M. (1985). Actinomycin and DNA transcription. *Proc Natl Acad Sci U S A* *82*, 5328-5331.
- Solovei, I., and Cremer, M. (2010). 3D-FISH on cultured cells combined with immunostaining. *Methods Mol Biol* *659*, 117-126.
- Solovei, I., Grasser, F., and Lanctot, C. (2007). FISH on Histological Sections. *CSH Protoc* *2007*, pdb prot4729.
- Solovei, I., Kreysing, M., Lanctot, C., Kosem, S., Peichl, L., Cremer, T., Guck, J., and Joffe, B. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* *137*, 356-368.
- Solovei, I., Thanisch, K., and Feodorova, Y. (2016). How to rule the nucleus: divide et impera. *Curr Opin Cell Biol* *40*, 47-59.
- Soutoglou, E., and Misteli, T. (2007). Mobility and immobility of chromatin in transcription and genome stability. *Curr Opin Genet Dev* *17*, 435-442.
- Steurer, B., Janssens, R.C., Geverts, B., Geijer, M.E., Wienholz, F., Theil, A.F., Chang, J., Dealy, S., Pothof, J., van Cappellen, W.A., *et al.* (2018). Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc Natl Acad Sci U S A* *115*, E4368-E4376.

- Stevens, B.J., and Swift, H. (1966). RNA transport from nucleus to cytoplasm in *Chironomus* salivary glands. *J Cell Biol* *31*, 55-77.
- Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* *332*, 472-474.
- Takizawa, T., Meaburn, K.J., and Misteli, T. (2008). The meaning of gene positioning. *Cell* *135*, 9-13.
- Tan, R.Z., and van Oudenaarden, A. (2010). Transcript counting in single cells reveals dynamics of rDNA transcription. *Mol Syst Biol* *6*, 358.
- Tanabe, H., Habermann, F.A., Solovei, I., Cremer, M., and Cremer, T. (2002). Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mutat Res* *504*, 37-45.
- Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* *159*, 635-646.
- Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.M., Robert, M.C., Capozzi, S., Chouaib, R., Backer, V., Mateos-Langerak, J., *et al.* (2016). A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat Commun* *7*, 12248.
- Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* *10*, 1453-1465.
- Trajkovic-Arsic, M., Muller, J., Darras, V.M., Groba, C., Lee, S., Weih, D., Bauer, K., Visser, T.J., and Heuer, H. (2010). Impact of monocarboxylate transporter-8 deficiency on the hypothalamus-pituitary-thyroid axis in mice. *Endocrinology* *151*, 5053-5062.
- Trask, B., Pinkel, D., and van den Engh, G. (1989). The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* *5*, 710-717.
- Trask, D.K., and Muller, M.T. (1988). Stabilization of type I topoisomerase-DNA covalent complexes by actinomycin D. *Proc Natl Acad Sci U S A* *85*, 1417-1421.
- Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* *28*, 1-13.
- 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.
- Van Sande, J., Raspe, E., Perret, J., Lejeune, C., Maenhaut, C., Vassart, G., and Dumont, J.E. (1990). Thyrotropin activates both the cyclic AMP and the PIP2 cascades in CHO cells expressing the human cDNA of TSH receptor. *Mol Cell Endocrinol* *74*, R1-6.
- van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* *169*, 780-791.
- Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* *10*, 1297-1309.

Volpi, E.V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R.D., Goldsworthy, M., Powis, S.H., Ragoussis, J., *et al.* (2000). 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* (Pt 9), 1565-1576.

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

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

Weiss, A., and Leinwand, L.A. (1996). The mammalian myosin heavy chain gene family. *Annu Rev Cell Dev Biol* *12*, 417-439.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* *153*, 307-319.

Williamson, I., Lettice, L.A., Hill, R.E., and Bickmore, W.A. (2016). Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* *143*, 2994-3001.

Wong, J.J., Au, A.Y., Ritchie, W., and Rasko, J.E. (2016). Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. *Bioessays* *38*, 41-49.

Wu, B., Eliscovich, C., Yoon, Y.J., and Singer, R.H. (2016). Translation dynamics of single mRNAs in live cells and neurons. *Science* *352*, 1430-1435.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* *15*, 1263-1271.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* *21*, 1273-1283.

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

Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* *148*, 908-921.

Zhang, Y., Yin, C., Zhang, T., Li, F., Yang, W., Kaminski, R., Fagan, P.R., Putatunda, R., Young, W.B., Khalili, K., *et al.* (2015). CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. *Sci Rep* *5*, 16277.

Zheng, H., and Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol* *20*, 535-550.

Zhu, L., Zhang, Y., Zhang, W., Yang, S., Chen, J.Q., and Tian, D. (2009). Patterns of exon-intron architecture variation of genes in eukaryotic genomes. *BMC Genomics* *10*, 47.

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

Zwemer, L.M., Zak, A., Thompson, B.R., Kirby, A., Daly, M.J., Chess, A., and Gimelbrant, A.A. (2012). Autosomal monoallelic expression in the mouse. *Genome Biol* *13*, R10.

## 5.2 SUPPLEMENTARY TABLES

TABLE S1: Primer sequences to introduce protospacer sequences into U6-gRNA-GFP-H2A. Overhangs for circular amplification are indicated (gray).

gRNA name	Target region upstream of TSS	gRNA sequence (5' → 3')	PAM (5' → 3')	Forward primer (5' → 3')	Reverse primer (5' → 3')	Laboratory internal plasmid identifier
<b>Ttn-1</b>	-52 bp	GAGCCGGGCTGT AAGGATGT	TGG	GAGCCGGGCTGTAAAGGATG Tgttttagagctagaataagcaagtt	ACATCCTTACAGCCCCGGCT Cgggttttcgctctttccac	pc3622
<b>Ttn-2</b>	-97 bp	AAATAGGTCCTCT CCCATCA	GGG	gAAATAGGTCCTCTCCCCATC Aggttttagagctagaataagcaagtt	TGATGGGAGAGGACCTAT TTcgggttttcgctctttccac	pc3623
<b>Ttn-3</b>	-143 bp	AATTAGCACTGC CAATCAG	AGG	gAATTAGCACTGCCAATCA Ggttttagagctagaataagcaagtt	CTGATTGGCAGTGCTAAA TTcgggttttcgctctttccac	pc3624
<b>Ttn-4</b>	-188 bp	AAAATACAACTCT ATTTCCT	CGG	gAAAATACAACTCTATTTCCT gttttagagctagaataagcaagtt	AGGAAATAGAAATGTATTT Tcgggttttcgctctttccac	pc3625
<b>Ttn-5</b>	-168 bp	TTTCAGAAAGAGG CTTACCG	AGG	gTTTCAGAAAGAGGCTTTTACC Ggttttagagctagaataagcaagtt	CGGTAAAGGCTCTCTCTGAA Aeggttttcgctctttccac	pc3836
<b>Ttn-6</b>	-261 bp	ACTTACCCAGGT GGAACGG	AGG	gACTTACCCAGGTGGAAACG Ggttttagagctagaataagcaagtt	CCGTTCACCTGGGTAGA GTcgggttttcgctctttccac	pc3837
<b>Neb-1</b>	-23 bp	GACCATCATGAGA GCTGGC	AGG	GACCATCATGAGAGCTGCG Cgttttagagctagaataagcaagtt	GCGCAGCTCTCATGTATGG TCgggttttcgctctttccac	pc3631
<b>Neb-3</b>	-240 bp	GCTGTTGAGAAAG AATCCGAC	CGG	GCTGTTGAGAAAGAAATCCGA Cgttttagagctagaataagcaagtt	GTCCGGATCTCTCTCAACAG Cgggttttcgctctttccac	pc3633
<b>Neb-4</b>	-343 bp	GAAAATAAGCCAC GACCATC	GGG	GAAAATAAGCCACGACCATC gttttagagctagaataagcaagtt	GATGGTCGTGGCTTATTT Cgggttttcgctctttccac	pc3634
<b>Neb-5</b>	-442 bp	AGACTACTGATGT GAGGAGC	AGG	gAGACTACTGATGTGAGGA GCgttttagagctagaataagcaagtt	GCTCTCACATCAGTAGTC Tcgggttttcgctctttccac	pc3838
<b>Neb-6</b>	-554 bp	TTCAGACCTCACT GAACGCA	TGG	gTTCAGACCTCACTGAACGC Aggttttagagctagaataagcaagtt	TGCGTTCAGTGAGGTCTG AAcgggttttcgctctttccac	pc3839
<b>Neb-7</b>	-494 bp	AGTTATTCATGGT CCCATCA	TGG	gAGTTATTCATGGTCCCATC Aggttttagagctagaataagcaagtt	TGATGGGACCATGAATAA CTcgggttttcgctctttccac	pc3841

TABLE S2: Overview of BAC clones used in this study.

BAC identifier	Laboratory internal identifier	Location	Organism
RP24-229C15	BAC 92	<i>Tg</i> gene body	<i>Mus musculus</i>
RP23-193A18	BAC 98	<i>Tg</i> 5' half	<i>Mus musculus</i>
RP23-266I10	BAC 99	<i>Tg</i> 3' half	<i>Mus musculus</i>
RP24-171I15	BAC 223	<i>Tg</i> 5' flank	<i>Mus musculus</i>
RP24-312F21	BAC 224	<i>Tg</i> 3' flank	<i>Mus musculus</i>
RP11-806J1	BAC 18 (human)	<i>Tg</i> gene body	<i>Homo sapiens</i>
CH216-26O19	XTR 2	<i>Tg</i> gene body	<i>Xenopus tropicalis</i>
CH211-184D12	DR 2	<i>Tg</i> gene body	<i>Danio rerio</i>
CH261-128F8	CH 2	<i>Tg</i> gene body	<i>Gallus sp.</i>
RP23-310F9	BAC 162	<i>Ttn</i> gene body	<i>Mus musculus</i>
RP23-140P12	BAC 234	<i>Ttn</i> 5' half	<i>Mus musculus</i>
RP23-314B24	BAC 237	<i>Ttn</i> 3' half	<i>Mus musculus</i>
RP23-1M1	BAC 238	<i>Ttn</i> 5' flank	<i>Mus musculus</i>
RP23-316G15	BAC 231	<i>Ttn</i> 3' flank	<i>Mus musculus</i>
RP23-82J15	BAC 176	<i>Neb</i> gene body	<i>Mus musculus</i>
RP23-91F4	BAC 257	<i>Neb</i> 5' half	<i>Mus musculus</i>
RP23-177B19	BAC 256	<i>Neb</i> 3' half	<i>Mus musculus</i>
RP24-158B9	BAC 225	<i>Myh11</i> gene body	<i>Mus musculus</i>
RP24-283P15	BAC 229	<i>Myh11</i> 5' half	<i>Mus musculus</i>
RP23-291J21	BAC 230	<i>Myh11</i> 3' half	<i>Mus musculus</i>
RP23-265A17	BAC 244	<i>Cald1</i> 5' half *	<i>Mus musculus</i>
RP23-386A17	BAC 243	<i>Cald1</i> 3' half *	<i>Mus musculus</i>
RP23-39F15	BAC 143	<i>Dmd</i> gene body **	<i>Mus musculus</i>
RP23-35L6	BAC 144	<i>Dmd</i> gene body **	<i>Mus musculus</i>
RP23-191F11	BAC 145	<i>Dmd</i> gene body **	<i>Mus musculus</i>
RP23-386E22	BAC 146	<i>Dmd</i> gene body **	<i>Mus musculus</i>
RP23-121C23	BAC 147	<i>Dmd</i> gene body **	<i>Mus musculus</i>
RP24-70A8	BAC 148	<i>Dmd</i> gene body **	<i>Mus musculus</i>
RP23-427K5	BAC 174	<i>Acta1</i> gene body	<i>Mus musculus</i>
RP24-332J21	BAC 111	<i>Sry</i> gene body	<i>Mus musculus</i>

\*For visualization of the entire *Cald1* gene, BACs RP23-265A17 and RP23-386A17 were combined.

\*\*For visualization of the entire *Dmd* gene, BACs RP23-39F15, RP23-35L6, RP23-191F11, RP23-386E22, RP23-121C23 and RP24-70A8 were combined.



TABLE S3: Primers to amplify *Tg* exons

Probe	Sequence (5' → 3')
Ex 2 – 12	fwd: AGGTAGATGCACAGCCACTC (in exon 2) rev: CAGGTTCCCAGCCTCCAATC (in exon 12)
Ex 33 – 47	fwd: TGCTGCACTGGCTTTGGTTT (in exon 33) rev: AGGAGCAGTCAGCTTGTTTGA (in exon 47)

**TABLE S4: 96 primer pairs to amplify probes for the start of *Tg*.** The probe covers a region of 32,404 bp. Since the fragments amplified by PCR are not directly adjacent to one another but are interspersed by stretches of 10 to 587 bp, the effective probe size is 20,021 bp. Lengths of the respective amplicons are indicated on the right.

	Chr.	Start	End	Forward primer (5' → 3')	Reverse primer (5' → 3')	Length (bp)
1	15	66670771	66670984	TTCATCCCAGTAGGGGACAG	CTCTGGGCATTTGTGAGGTT	213
2	15	66671026	66671228	GAGAGATGAGCTGAATCGCC	CCCTGACACAGACACCTTT	202
3	15	66671247	66671451	GTAGGACTGGGAGGAAAGGG	AGGACCCTGGAGCTCCTAAT	204
4	15	66671529	66671742	TATGTTGTCATTTTCCGGGG	TGCCCCTACCAGAGAATTGG	213
5	15	66671828	66672035	TGAAAGGAGCTTTTGCAGGT	TGGGCCTATAGGAAGCTTGA	207
6	15	66672233	66672434	CCTCTCCTGCCTTGTGTGAT	CTGCCAGGTACTTCTCTGCC	201
7	15	66672501	66672707	ATACTGGGGCACCTCACATC	CTCAGTGTCTGGACACAGTGG	206
8	15	66672795	66673004	GGATGCTCTTTTTCAGTGGG	ATTGCAAACTCTAACGTGGG	209
9	15	66673127	66673329	GAAGCAACTTTTCAGGCCAC	GAAGGATTGCAGGATCAGGA	202
10	15	66673526	66673729	CACAGAGGGGATGGAAGTGT	TGTGGGATGACCAAAAGTCA	203
11	15	66673957	66674159	TAGGAGACAGGTACACACCC	AACTCCAGATCCTGATGC	202
12	15	66674924	66675130	TGGCTGTGTTGTTCTCCATC	AAITCTACCCITGCACTGCC	206
13	15	66675384	66675590	TCCTGTCAAAGGGGAAAGAA	TCATTTAAAGGCTGCATGGC	206
14	15	66675846	66676048	AGCCTGGCGTTTTCATTAGA	CAGGGACCTAGGACACAGGA	202
15	15	66676091	66676308	GTACAATCCTGCCCCATAGC	CAGATGATCACGTTGGGATG	217
16	15	66676364	66676576	CCACCTGGCGACTACAGATT	TTCCCAGGGTGAGCTCTTTA	212
17	15	66677065	66677284	TTTGAACCACTGGAAGGAGC	TAAAGTGCAGGCATCTGGTG	219
18	15	66677390	66677605	GCACCTAGGAGGTGGAGCTG	GGGATTGGGAAGGACTGACT	215
19	15	66677762	66677963	CCCCTATATGTGGGGACAAG	AGCAGAGTCCTCAGCTGCAT	201
20	15	66678005	66678205	GTGACCACGCATCTTCCTCT	GCCTTAGGCACCAGAGTCAA	200
21	15	66678304	66678517	TCTGCTTGCTTGTATGGTGC	TAAGCTAGATGCAGCCCCTG	213
22	15	66678560	66678778	TGTTGTGTGCTTTTGCATTG	TTCCAGTCTGAGCAAAATGT	218
23	15	66678856	66679060	CCGGATATTGCAGAGACGAT	TGTTACAGCCTGCATGAAT	204
24	15	66679215	66679429	GACAGAAGGGATGTGCTGGT	CAGTTTGAAGGACCTCTGGG	214
25	15	66679653	66679854	CTTCATGGGTGAGGGACAGT	CATATCAGCCTACCACCCT	201
26	15	66679885	66680089	CATGTGGGATATACGCCTCC	AGTTCCTGCTTTGACTCGGA	204
27	15	66680304	66680509	ACGCACACACACACACACAC	CGTAAGCAAGCTCCATCCTC	205
28	15	66680653	66680858	CACAAGGGGTGCTACAACAA	CTCTGGTTCCTCTGGTGTCC	205
29	15	66680956	66681156	TGTTAGCAAACCTGGGCTCCT	ATAGATGCAGCCATGAAGCC	200
30	15	66681190	66681404	CCTCTTGTCTGTGGCTCTTC	GCCTATGAGATACCAGCCCA	214
31	15	66681823	66682038	TCAGTCATGTGCCITGGAAA	GAAAGCCGTTGGTAGTGCTC	215
32	15	66682109	66682321	CTTCAGTTCACCACCAACCC	GGAGTCCAATCTCACAGGGA	212
33	15	66682381	66682591	TGTGGGTAACCTTTGGCTTCA	TGGGACAAAGAACTTCCCAG	210
34	15	66682676	66682880	GATGGCTCGAGAGTCAGAGG	CAGGAATAACCTGGCCTTCA	204
35	15	66682897	66683108	TGGAGAAGCCAAGCAGTGTA	CAAGCAAATATCTTTATCTGCTCATT	211
36	15	66683127	66683326	TACATACCTGGGCTGGGAAG	AGCCTGGAGCTTTCCAATTA	199
37	15	66683661	66683880	CTGAACAAGCTTTTCTTGGG	ATCTTCATTAGCAGTGCGGC	219
38	15	66683900	66684112	GCTTCCAGAAACTTCAGCCT	GTAGGGACCTGGGTATTGGC	212
39	15	66684154	66684373	CTGTTGGTGTGTGGATGAGG	GGGTGGATAGGTGTGAATGG	219
40	15	66684953	66685163	GGCTGTGGCCTAAAATGGTA	GGAAACCTGCAGAAGACTGG	210
41	15	66685389	66685590	GACCAGTGAGGAGCTTGACC	TGCTCATGGAAACCAATGA	201
42	15	66685653	66685853	CATGATCCTTCAGCAAAGCC	AAAGGAGCAATGGGTCTCTCA	200
43	15	66685874	66686090	GCAGGTCCCAAAGGTTTGTA	GGAGTAAACCTGGGCAAGTG	216
44	15	66686301	66686516	CTCTGAGCCTTGTCTGAGCC	GAGTGGCAGCAACAGACTGA	215
45	15	66686732	66686950	CACAAAATGCAGAGGAGCTG	ACCAAAATCCCCAAACTTCC	218
46	15	66686979	66687198	CAAAGCACATTGGCTGGTA	GGTACCAATCTCTTGGCTG	219
47	15	66687225	66687428	TTGATAGCTCCATGCACAGC	TAACCAGGTGTGGCCTTTC	203
48	15	66687520	66687727	AACTGGGAGACCCAGACTCA	TGTGGCTCAGTGGAATAGGA	207
49	15	66688109	66688311	GCCATGCAAGGATAGGAGAG	TTTCAGAGCTGCTGACTGGA	202
50	15	66688339	66688558	GGACCAACACAAAACCATAA	GGTAAATGACTTCACGCCCT	219

51	15	66688577	66688788	TGAAATTACCACGCAGAGCA	CCACGGAGAGAAAAAGGACA	211
52	15	66688846	66689052	CAGCCAGTCTTCTGTGGTCA	TAGTTGAGAAGGCAAGGGCT	206
53	15	66689102	66689319	GAAATTCCCATTTGCTGGAA	CGAGCCTGGAATGAACTCTC	217
54	15	66689415	66689622	GCTTGGTTCATGCAACTGAGA	GCTAGGAGCCAGAACACAGG	207
55	15	66689739	66689946	CCCACATCTGTCCCTGAACT	GCCTGTTTCCATGCAGAAAT	207
56	15	66689969	66690180	CCGGGAGATCTGTTTCATACC	ACTATCCAGTGGCCCACATC	211
57	15	66690396	66690610	TCTGCTGGCAGTTCATTGAC	TCGTGATAGCCACACCACAT	214
58	15	66690678	66690885	TTGGTTGACAAACCACCTTT	AGCAGTTTGGGTAGTGAGGG	207
59	15	66690899	66691113	GCCTCCCTTTACCTCCAAAG	ATCTGCTGCCCTCATCATTC	214
60	15	66691125	66691335	CCTTCTGAGTGTCTGTGGA	TTTGGAATTAAGTCCCTCG	210
61	15	66691786	66691988	AAACCCTGCTGACACAGGAC	TCCACACTGACCAGAACCAG	202
62	15	66692090	66692289	GGCAAAACTTGTCTTGTCTT	CCCGTCTGCAATAAAACCAT	199
63	15	66692318	66692527	ACAGGAACATGGTGTGTGGA	GCCTGTGCTCTAGATGGGAA	209
64	15	66692560	66692759	CAGTATGGCCAGGAGTCTGC	GTAAAGAGCCCAGGGTTGGT	199
65	15	66692925	66693125	ATCTACCTTTCCCCACGCTT	TGTTTCTCCCATTCGCTACC	200
66	15	66693150	66693349	TGCCTAGTAGGGCAGGGGTA	ACTGGGGAGAAGGCTCCAT	199
67	15	66693371	66693574	CAGTTGTTGGTGTGTCTTGG	GTTTCCCACCACATCAATC	203
68	15	66693678	66693891	GTCCCAATCATGTGATGC	CACCAAAATGGGACTCAAAC	213
69	15	66694061	66694260	ACCTTGATGACTGAGGTGTCC	TCAGTCACTAAGAGTTCCCCAAA	199
70	15	66694293	66694492	TCAGAGAATCCCGAAGTGCT	CATCCTCTGTACCTCAGGGC	199
71	15	66694621	66694838	CCTGACAGAGCAGGACACAA	TGCTGAACAGTCGTCATCC	217
72	15	66695021	66695221	ACAGCCTTTTCACGTCAGGT	GTCCCTAACTCAGCAGCAGG	200
73	15	66695264	66695483	CCCTAACCTCTCCCCACATT	GGCTATGCTATGTCCCATGC	219
74	15	66695540	66695741	CTTCTGGGGTCAGTGGTAG	AGCTGGGACCACTCTGTCTAT	201
75	15	66695849	66696066	ATGGGAGACTTTCCAGGGAG	TGGGGCAATAGGTCTGTAGG	217
76	15	66696114	66696320	CCATGCAAACTCAAGCACAC	AATCTGAGGCTCCCTGACAA	206
77	15	66696341	66696550	CCACTCTTCAGGAGCAGGAC	AAACGCTCTGCAGTCAGACA	209
78	15	66696632	66696835	TACCTGAACTGCTGCCTTGA	AGTGGCTCACAACCACCTGT	203
79	15	66697291	66697498	GGTGGAGAAAGGAGTCTTCA	CCACCCTACCAGAAAGCTCA	207
80	15	66697562	66697775	GTGGCCAGACACCCAGTAGT	AGAGGAAGGCTGACCCAAAT	213
81	15	66697829	66698038	TGTCCCAGCTTCAGTGTGAG	GCATACCCTTCTGTTGGGAA	209
82	15	66698161	66698366	ACCTGAATAGTGCACCCAG	GCTGCCTAGAATGGCTTGAC	205
83	15	66698429	66698646	CTGTCTTACAGAAAGGGCGG	GAGCATCCTGTCTGGTGGTT	217
84	15	66699233	66699436	CATGGGTCTCCTTGGCTAAA	GCTGTGGAAGTTTCTGGAG	203
85	15	66699526	66699738	GATGTGGTGGTGTGTGGGTA	CAAATTCACITGGCAGGGAT	212
86	15	66699753	66699961	CCCAGCACAGAGTACAGAAGC	TATGGTGACTTCAAGCCCT	208
87	15	66700100	66700302	CCCAGAGAGGGCATCTCATA	CTAACCCCAATCTCCTGGT	202
88	15	66700393	66700607	AAAATCCAGAAATGGTGCCTG	GTAGAGCAAATGAGGTGGC	214
89	15	66700617	66700831	TTGAGCAGCTTCACTTCAGC	AAATTCAGTGGGGCTTCTCA	214
90	15	66700897	66701103	AGCAAGCTTTGGAAGCTGAG	TCAGGGGTACTCACTGAGGG	206
91	15	66701232	66701449	GGCTTCTCCTTCTTGGCTTT	GGGTGAGATGACCATGTCC	217
92	15	66701491	66701709	GCTGTGGTTTTCTCAGGACA	TCCACAGGGAACGAGTTCTA	218
93	15	66701752	66701967	TTCTCTCCCTTACCCCTTTTC	GCTCCTTTTTCAGACCCCA	215
94	15	66702322	66702532	GGAGGTGACCCATCAAGAAA	CCCAGCAGTGACAGGTCATA	210
95	15	66702709	66702920	CTGGTCTTGAGAGGAGCCAC	GTCCCGATGAAGACAGGAAA	211
96	15	66702962	66703175	CATTAGTTTCTTTGGCCCC	CTTGGTCTGCATGCATTCTT	213

**TABLE S5: 71 primer pairs to amplify probes for the first half of *Tg* Intron 41.** The probe covers a region of 24,622 bp. Since the fragments amplified by PCR are not directly adjacent to one another but are interspersed by stretches of 6 to 825 bp, the effective probe size is 14,779 bp. Lengths of the respective amplicons are indicated on the right.

	Chr.	Start	End	Forward primer (5' → 3')	Reverse primer (5' → 3')	Length (bp)
1	15	66773879	66774078	TCAGGAGCTATTCTGCCACA	GGGGCCTCTTTTAAACGTC	199
2	15	66774149	66774363	AAGGCTGTGGAAGCATTGAC	GCAAGACCAAGTTCTGGAATG	214
3	15	66774390	66774591	CTCAGGGATGGAGCTGTCTC	AAGTACAGGGCTCCCATTC	201
4	15	66774827	66775044	CCCATTTATTTTCCTTTTGGC	TCATTTCCCAAATACCTCG	217
5	15	66775363	66775566	ACATGCACATGCACACACAG	CATCAAGCATCAGTTTGGC	203
6	15	66775709	66775911	CCCTGCCCATTCATCATTTA	ACCTGGGGGAAAAGGTACAC	202
7	15	66775963	66776173	CAGCCAGTGTGCAAGTGAAT	CCAGGATCAGCCTAGAGCAC	210
8	15	66776193	66776406	TGTTAGCAACCAGACCCCTT	AAATAAGCTCACCCGATCCC	213
9	15	66776439	66776658	TCATGGCTCTTCAGTGTTC	CCAGTCAGTGGGAAGCATCT	219
10	15	66777183	66777389	TCCAATTCCAAAGGAATCCA	TCAGCTGAATTCCAGAGGCT	206
11	15	66777502	66777720	TAGCTATGGAAATGGCCCAG	CAGGCAGAATTCTCAAAGC	218
12	15	66777943	66778143	TTCCGCTAACAAGAGTGATTCTC	CACCTCCACTATGTACAGGCA	200
13	15	66778178	66778396	GGCCAGTCTTGAGCACAAT	TGAAACAATCCAAGGACACTCA	218
14	15	66778464	66778665	ATTGGTTGGTGTGAATGGC	GCTTTGGTGTGGGTCTCTA	201
15	15	66779279	66779495	CACCCTCAAGACTCACTGGA	CTGGGTGCTGGAGAGCTAAG	216
16	15	66779715	66779925	GGCCACAATGCACCTTTAAT	TTGATGAACGTGCGGAAATA	210
17	15	66780145	66780359	TCAACTTCCACCTAGCGTCC	CCTGTAGGCCTGAGCAAGTC	214
18	15	66780509	66780708	GTTAATGGCATCTCTCTCCA	GAGAGAAGGGATCAAGAGGGA	199
19	15	66780731	66780935	GGGACCCTGAGTCAGAACAG	GGGAAGCATTACACTTTGCC	204
20	15	66780979	66781196	TGACCACAACAGAAAACAGCA	GCCTTGACATGACGTGACATA	217
21	15	66781236	66781453	GAATGCACATGATGGGTGAC	GGAAAGGGGAACTGGAGAAG	217
22	15	66781632	66781832	TCCATGGGAATAAGAGCACC	TGCCTGGGTGAAATACTGTG	200
23	15	66782026	66782228	TGCTAAAGATTCCATTCGGC	AGACCATGTGGAAAGCATCC	202
24	15	66782245	66782450	GGATTTGAGTTGGAGTCAGTC	GGTCTTCCAGGAATCAAGGG	205
25	15	66782486	66782698	ACCCAGGTGACAGTCAAAG	CAGCTATGGTCTTCGGGAAA	212
26	15	66782717	66782927	CTGAGTGGGTTCCTGTCTCC	CCACCATTGCAAGAAATGAA	210
27	15	66783109	66783322	AGCCACATATACACCTGG	TCACCCTGCTCTTCTGATT	213
28	15	66783685	66783887	GGGATGTTGGAGCAGGTATG	CCTTTAGCATCTTGGCCTCA	202
29	15	66783999	66784212	TTGCTTTGATCTGATGTGTC	ACATATGCTTTGGAGGTGCC	213
30	15	66784332	66784531	AAATAGGACGGGGAGGAGAA	GGGAGTCATATCCTGTGCT	199
31	15	66784677	66784883	CCAGCAGGACACGGTTATCT	GTGGTGTGTCTCCCTCACT	206
32	15	66784999	66785217	AGTCAACATGACCCAGAGGC	ACAGGCAGGACAGGGTTATG	218
33	15	66785233	66785441	ATCTCATTCACCCTGGTCCC	AAGCTGGTGCAGTGTCTCAA	208
34	15	66785457	66785670	GTCCAGAAGCTCAGGTGGAG	AAAGAGGAAGGGAGTCTGG	213
35	15	66785736	66785949	TAGACATGGCACAAGGACCA	AAAGGGCATTTCAAGGAGGT	213
36	15	66785996	66786202	TCCAGTACAAGGTCAAGCA	GAGGCACCGTGAGAATTGT	206
37	15	66786286	66786502	CCACCCGGTAGACATGAGAT	TGGGGAGAATCACCAAGAAG	216
38	15	66786662	66786879	GGGGAGAGATTGACATTGGA	TGAGCATCCTGCTTGATAG	217
39	15	66786945	66787150	CAATGGCTTTTCTCAGGGAC	TGAGACACAGGCAGGTGAAG	205
40	15	66787294	66787506	AGGGGATCTTGCCCTATTG	AAACTGAGGAACAGGGAGCA	212
41	15	66787623	66787834	TACCATCACCTCTTGAGGCC	TCAGGAATCTGAAACCGAGG	211
42	15	66787861	66788064	TTGCTCTTCAGTCTAGTTTGG	CAGGGTGTACAGTTCATGC	203
43	15	66788088	66788304	TCAACCTTGGATCCTTTGAG	CTGTGAGTTTTCACAGGACC	216
44	15	66789079	66789278	CCAGGCTCATAGAGAGGCAG	AGTTACTTCTGGGCTGGGGT	199
45	15	66789436	66789643	GTATGTCAAGATGCTGCCCA	AACATATGGCCACATAGCCC	207
46	15	66789766	66789978	CCAGTCTTCAAACAGCCAG	ACTGGGCACAAATGGCTACT	212
47	15	66790037	66790251	TGGCACTGACTAGAGCAGGA	TCAGGAGCAGAACCACACTG	214
48	15	66790387	66790593	GGGAGGACCACTTTCACAGA	CTGAGTTAGGCCTGGGGATT	206
49	15	66790726	66790933	TTCCATTACCTGTGGCTCC	CCCCACATGGGTCTTGATAC	207
50	15	66791001	66791201	TACACAGGCAAACTTCCCC	ATAATCCTGGCCCTATCGGT	200

51	15	66791336	66791536	TGGACATGACTCCCATAGCA	CCTGAGCTAGGTCATGCACA	200
52	15	66791610	66791810	ATGACAGAAGGGAAGCCTGA	GCTGAAGTGGAGGTTCAGA	200
53	15	66791890	66792107	CTTGCCCGTTATCTCATGGT	ACCCCATCCTTGGGGTAATA	217
54	15	66792123	66792335	AGCAATGCTTTCTTCCACTG	TCCITTTCTGCTTTGTCTGA	212
55	15	66793160	66793373	CAAGTTACGTGTTACAGGGG	ACACGCTAAGGGCTCTGTGT	213
56	15	66793515	66793714	CACTGGAGGCTGTAAGGTCC	TCCGGAGAGGAGAGAACTG	199
57	15	66794359	66794565	GAACCTTACCCAGAAAGCT	CCTCCCTCCCTCCATCTATC	206
58	15	66794608	66794815	GAGAGAGAGAGAGAGAGAAACAGA	GCCTGACACTTTCCTCCAAG	207
59	15	66795238	66795437	AGCAACATTTGCTCCCATTC	GCCCAATGAGGAGTCACTA	199
60	15	66795533	66795747	TGCCTTGGCTTCAGAAAGTT	GCITCATTACCAATCAGCA	214
61	15	66795773	66795980	GCACACGCAAAATGTGTGTT	GGGCAGTGTTAGAATCATTAG	207
62	15	66796004	66796220	CTCTTCTCTTTGTGGCCAG	TATCTTGGGCTCCTTGATGG	216
63	15	66796252	66796452	GGCTGCCATTCAAGAGTCTG	AAAGACCTGTCGCTTCAGC	200
64	15	66796474	66796680	ATGAGCTGTCCCTCTGCCTA	CGTGCTCAGTTTCTAGGGG	206
65	15	66796749	66796949	GGGAAACAAAATGGAGCTGA	AATGTGACCACATCAAGCCA	200
66	15	66797097	66797306	GATTCACAGTTGCAGAGGCA	CACTGATGTTCCCTGGAGGT	209
67	15	66797329	66797533	CTTTAGGGCCCCACTGTCTT	CGTTTGTCAAAATGTCTGCC	204
68	15	66797539	66797748	GCTACCTGTTTAGGCAGACTTATACAT	CTCTGTGACTCTCATTACTGACCA	209
69	15	66797780	66797991	CATTTGAAATCAGCTGGGCT	GAGCTGGTCTTGATTTGGGAA	211
70	15	66798027	66798231	GAGGCAATGTGAGGCTTTCT	AGGCCTGCTCTTCGAGCTA	204
71	15	66798297	66798500	AGCCCCATTGGTGTCTGTAG	TGAGGAGAGCCTAGACCCAA	203

**TABLE S6: 74 primer pairs to amplify probes for the second half of *Tg* Intron 41.** The probe covers a region of 24,460 bp. Since the fragments amplified by PCR are not directly adjacent to one another but are interspersed by stretches of 8 to 622 bp, the effective probe size is 15,429 bp. Lengths of the respective amplicons are indicated on the right.

	Chr.	Start	End	Forward primer (5' → 3')	Reverse primer (5' → 3')	Length (bp)
1	15	66802529	66802730	TTTCCCATTTCCAGGAACAA	TATGAGGGGCATTTGAGGAG	201
2	15	66802961	66803177	TGGCTTTTGTGCCACTGTTA	TGGTCTCAAGGCTGACTCCT	216
3	15	66803299	66803516	TTGCACCTATATTTCCCCCA	GGGAGAGGTTACTTCATTACCCA	217
4	15	66803827	66804031	GCACACAGAGACAAACCCAA	ATGTGGTGCAATGCTATGGA	204
5	15	66804182	66804400	GGTGGAGAGAGGTGCAAGAG	AGGCTAAAGGGCACACAGAA	218
6	15	66804451	66804650	AGAGAGAGCCTTGTACCCA	CTCTCAGAGATGCTGCCGA	199
7	15	66804696	66804905	CTTGGGTTCCTGTAGAAA	AGGCAAAAGACCTGCAGAGA	209
8	15	66804930	66805149	TGGTTCTCCATCAGTGCAAG	CCTTGTGTGACTTCCAGGGT	219
9	15	66805246	66805460	CACTCCTGTGGCCATCTTCT	TTGAAGCCTCAGAGCATGTG	214
10	15	66805674	66805893	GAAACCTAGGCCTGGAACC	CTAAGCCAGCTTGGGCTATG	219
11	15	66805913	66806126	GAGCCTGTGGAGTACTGGGA	GCACATAAGGCAAGTGTGGA	213
12	15	66806203	66806413	TTATTCAGGTCCCAATTGC	CACTCTGTACACCAAAAGGC	210
13	15	66806451	66806654	TTTCTGGGCCACATTACCTC	CTGACAATTGAGGCCAGGAT	203
14	15	66806677	66806876	AGCAAGACTACCAAGGCAGC	TGGCTGTGAGTTAGGCACAC	199
15	15	66807040	66807253	TGGGCATTGCTTTACACAGT	CTCTCAGGCTCCTTCTCCTG	213
16	15	66807292	66807495	GAACCCCAACACTAGCAAGG	TGACGTGAGATCCTAACCAGG	203
17	15	66807514	66807723	GCTCTCCCAACAAGAGTCCAG	GCACCCATAGCTTTTGCTTT	209
18	15	66807860	66808073	GAACCTCCCAGCAGTACCAG	GGGCACAGAAATGGAGAAA	213
19	15	66808695	66808908	TTAGGCCACACTCAGGTATG	TGAAGTGGTGGAGCTCAGTG	213
20	15	66808965	66809178	GGCCTCCATTGTGCTTAAAA	CCTCTGTACCTTCTGCAGGG	213
21	15	66809196	66809412	CTGAAAGCCAATTGCACTG	GAGGGATAGTGTCTTGCCCA	216
22	15	66809442	66809653	GTAGGCAGAGCAGATGGAGG	TTCTGAAAAATGGCAAAGG	211
23	15	66809811	66810013	CCCCAGGGCAATAGATACAA	TGTACTCTGTTGGGGTTCC	202
24	15	66810193	66810393	TGGTCAGAAATGCTCCTTCC	TACCTATCATCCAGCCAGC	200
25	15	66810624	66810842	CCCCTCAGGTGACTAGCAGA	CTGGGGAAGGATGACAGAA	218
26	15	66810869	66811072	TGTTAGGAACACGTTCCACTTG	CAGGCAAAGCCTGCCTAAT	203
27	15	66811221	66811439	GTCTTTTGATGGCAAACCTG	AACTGAGCTGTGTGGCAGTG	218
28	15	66811525	66811724	CTTTCTTTCCCTGTGAAGGC	CCTGCTGTGACTGCATTAAA	199
29	15	66811765	66811969	TCTGAAGGTTTCAGCCAGAT	CTGAAGTAGGCTGCTGGGAG	204
30	15	66812094	66812300	CCTCACTGGGTACCTCCTGT	CCAGGGGAAGTCTGACTCAA	206
31	15	66812322	66812522	CACCTAGCACAGCACCTTGT	AGAGAACTGTGGCTGGGACA	200
32	15	66812553	66812772	CCTGGGAGTTGGCAGTAGAG	AGAAAACAGGAACCTGCCCT	219
33	15	66812796	66813008	CTCTACAAGGACGATGGGGA	GCAGTGAGCTATCCTCAGCC	212
34	15	66813075	66813276	GGAGTCCCTGAAACTGTGCT	AAGGATCACGCCTTGCTCTA	201
35	15	66813387	66813601	AAATGGATGAGACACCCAGG	GCCTAACCCCATCCTCTCTC	214
36	15	66813625	66813824	CTGGATGTGCACACATTGCT	GCACTACTGCCTGCCATTTT	199
37	15	66813854	66814053	CCAAAGTTGAAAACCCCTCA	TAGCTTCTGCCCCACTGTTT	199
38	15	66814169	66814369	TCAGTGAGGAAGAGCAAGCA	CAGCCATGTTAAGGGAAGGA	200
39	15	66814473	66814680	CACATTTGTGGTGTCTCT	CCTCTGCCATTGATGTTCTT	207
40	15	66814876	66815089	CCCTAGTCCACCGAGATTGA	GGAGTGACCCATTTTGTGCT	213
41	15	66815307	66815524	TGCACTGAGGTTTAGCAGGA	TGGCTTGAGTGTCTTTGTGG	217
42	15	66815646	66815846	GGCACTGTGGTTTCTACCGT	GCCTCACATATCTGCCATGA	200
43	15	66815941	66816156	AAGAACCAGGTAGCATCCCTT	TGTTGCAAAATACAGCTCCCA	215
44	15	66816164	66816383	CAAAGTCAAGAGACGAGGC	CCACCTCCTGTGCCATTAT	219
45	15	66816399	66816612	GCCACTGAACCTACCCTGAA	CTCAGATCTTCCACAGGCT	213
46	15	66816630	66816846	GTTTCTAAACCGCGTGTGT	TATGGAGCTGAGAACCACCC	216
47	15	66816876	66817079	AGCTTGATCGCCATGTCTT	TAGATGCGTTTTCTGGGGTC	203
48	15	66817138	66817345	CAGCTAGACATCTGAGGGTCTC	GGAGGGGATGCAGGATTATT	207
49	15	66817367	66817579	CCTGGGCACTTCATGGATAC	AAGATGGAAGCCAGGAGGAT	212
50	15	66817783	66817987	TTCATGTGGAGCTTCAGTGG	AAAAAGCTGTCTGTGAGGCG	204

51	15	66818148	66818358	CGGCTCACAGACAGCATTT	GGGGCCCTTCTTTCTCTCTG	210
52	15	66818434	66818633	CAGAAAGTTGCCGGAATTAGC	GCCAGTTACTTTGGCATGGT	199
53	15	66818699	66818903	CACTCACTGAAACCCACCT	ATGGATCAGTGCATGCTGAG	204
54	15	66819148	66819353	CTTGGTCTTAGGCAAGCCCT	TGTGAAAATCAAATGAAACCACA	205
55	15	66819411	66819619	GCTTGGTACACATAAGAAGCCC	AGAAGCATTGAGAGCATGGG	208
56	15	66820237	66820454	CAAACATGGAGGTGCAAAAG	GTGTTTCTAGATGGACTTGGCA	217
57	15	66820533	66820747	AGCTGCGGAGATGAAAAGAG	TGATGGGAAAATAGGCCTTG	214
58	15	66820777	66820996	GCAGGCTAGGTCTTGCATCT	TATGGGGAGAAGGGGACAG	219
59	15	66821017	66821236	GCTGGTCTCAGGAATGAATGA	CAAAACACCTTTTGCCATCTG	219
60	15	66821266	66821471	ATAGGTGCAAAATCCGTTCCA	GGTATGGCTAAGGCCATTGA	205
61	15	66821500	66821701	TGTGCAAAATGTCTGCTGTCC	CAATTCTTAAAGCTCCCTGC	201
62	15	66821734	66821933	GTGGGCAGGTAATAGGGTGA	TGTGTGTAAAATTTCCATAGATTGAGA	199
63	15	66822502	66822717	TATCATCTCCCGCTCTCTG	AGAAGGAGGAGGCAAGGTGT	215
64	15	66822828	66823033	AGGCCCTTATAAAGGCTGA	GTGTGGCTTGCTATTGCTCA	205
65	15	66823251	66823463	TGCTCTTGTTAAAGGCAT	CCAGAGAATGTGTCCAGTGC	212
66	15	66823785	66823984	TTTCAGTCTGGCCAAAATCC	ACAAACAGAAATGGATGGGGA	199
67	15	66824041	66824252	CCCAGGTTTTTCTTCAACA	GCTATTGGGACCAGGACAAA	211
68	15	66824441	66824642	GGGTGTTGCACTGCTGTAGA	ACCCTGTGATCAACTGCTCC	201
69	15	66824717	66824929	ACAGGGCCTTCTACTGGAT	TATCAGAACCGCCCATGAAT	212
70	15	66825435	66825637	CCCAAATGAGGACAGCCTAC	GGGCCTCTGATGTTTTCAAT	202
71	15	66825952	66826158	CTATGCCGCACTGTGAAGAC	TGCACATCAAACCTTGCATCC	206
72	15	66826295	66826509	AAACCCCTGTCTGCTACCT	ACCTGTGACCAGAGCCTGAG	214
73	15	66826564	66826765	ATCAGGAGCCTCCAGGTAT	AGCACCCAGAACTCAGCAGT	201
74	15	66826790	66826989	AGGCACCAGCAACTTCTCTG	TGATATAGCTCTGAATGCCGAC	199

TABLE S7: Primers used for qPCR

Gene	Sequence (5' → 3')	Organism
<i>Pax8</i>	fwd: CAGCAGTGGTCCTCGAAAGC rev: GGTGCGTCCCAGAGGTGTA	<i>Mus musculus</i>
<i>Tg</i>	fwd: CCATGGCAGCCTAGAATTGC rev: ACTCATGTGGGTAGTTAGGATTTCCT	<i>Mus musculus</i>
<i>Gapdh</i>	fwd: CATGGCCTTCCGTGTTCCCTA rev: CTTCAACCACCTTCTTGATGTCATC	<i>Mus musculus</i>
<i>Ttn</i>	fwd: GCCGCGCTAGATTGATGATC rev: TCTCGGCTGTCACAAGAAGCT	<i>Mus musculus</i>
<i>Neb</i>	fwd: CAGCAGATGCAGAGTGGGAA rev: TAACAATGCTGGCGTGACCT	<i>Mus musculus</i>



TABLE S8: Long control genes with low expression used for depiction of the ratio of *cis*/ *trans* contact frequencies.

Gene	Expression level (TPM)	Length (bp)
<b>Myoblasts</b>		
<i>Pde4d</i>	482.09	1,503,513
<i>Magi2</i>	1.48	1,477,756
<i>Agbl4</i>	4.09	1,266,663
<i>Tenm2</i>	1.21	1,229,308
<i>Csmd3</i>	2.10	1,211,426
<i>Ccser1</i>	0.56	1,202,541
<i>Prkg1</i>	5.87	1,197,482
<i>Dab1</i>	7.281	1,125,485
<i>Aut2</i>	44.26	1,106,011
<i>Sgcd</i>	4.60	1,092,716
<b>Myotubes</b>		
<i>Dlg2</i>	2.06	1,358,541
<i>Agbl4</i>	2.07	1,266,663
<i>Park2</i>	0.39	1,222,977
<i>Csmd3</i>	0.62	1,211,426
<i>Prkg1</i>	13.82	1,197,482
<i>Dab1</i>	2.70	1,125,485
<i>Aut2</i>	32.30	1,106,011
<i>Sgcd</i>	10.71	1,092,716
<i>Nrg1</i>	0.55	1,069,478
<i>Rbms3</i>	21.95	1,057,167
<b>Thyroid</b>		
<i>Fhit</i>	15.57	1,611,943
<i>Pde4d</i>	236.00	1,503,513
<i>Magi2</i>	4.20	1,477,756
<i>Agbl4</i>	24.44	1,266,663
<i>Lingo2</i>	0.07	1,245,100
<i>Ccser1</i>	5.16	1,202,541
<i>Prkg1</i>	0.15	1,197,482
<i>Ptpn1</i>	0.29	1,139,157
<i>Knip4</i>	0.15	1,135,405
<b>Bladder</b>		
<i>Fhit</i>	10.45	1,611,943
<i>Pde4d</i>	379.15	1,503,513
<i>Magi2</i>	2.67	1,477,756
<i>Lrrc4c</i>	0.38	1,313,497
<i>Agbl4</i>	18.93	1,266,663
<i>Tenm2</i>	16.53	1,229,308
<i>Csmd3</i>	3.56	1,211,426
<i>Ccser1</i>	1.30	1,202,541
<i>Prkg1</i>	4.44	1,197,482

## 5.3 ABBREVIATIONS

3D	Three dimensional
Ac	Acetylation
<i>Acta1</i> / <i>Acta1</i> /ACTA1	Actin alpha 1
<i>Actb</i> / <i>Actb</i> /ACTB	Actin beta
BAC	Bacterial artificial chromosome
bp	Base pair
<i>Cald1</i> / <i>Cald1</i> /CALD1	Caldesmon
Cas	CRISPR associated
CC	Chromocenter
CDK	Cyclin dependent kinase
cHC	Constitutive heterochromatin
ChIP	Chromatin immunoprecipitation
cDNA	Complementary DNA
CRISPR	Clustered regularly interspaced short palindromic repeats
CT	Chromosome territory
CTCF	CCCTC-binding factor
CTD	C-terminal domain (of RNA polymerase II)
DamID	DNA adenine methyltransferase identification
DAPI	4',6-diamidino-2-phenylindole
<i>Dmd</i> / <i>Dmd</i> /DMD	Dystrophin
DNA	Deoxyribonucleic acid
DRB	5,6-dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole
DSIF	DRB Sensitivity Inducing Factor
EC	Euchromatin
EJC	Exon junction complex
EM	Electron microscope/microscopy
ESC	Embryonic stem cell
fHC	Facultative heterochromatin
FISH	Fluorescence in situ Hybridization
GFP	Green fluorescent protein
gRNA	Guide RNA
GTE <sub>x</sub>	Genotype Tissue Expression
HC	Heterochromatin
HIV	Human immunodeficiency virus
hnRNP <sub>C</sub>	Heterogeneous nuclear ribonucleoprotein C1

---

HP1	Heterochromatin protein 1
HSA	Human ( <i>Homo sapiens</i> ) chromosome
kb	Kilo base (pair)
LAD	Lamina associated domain
LBC	Lampbrush chromosome
LLPS	Liquid-liquid phase separation
Mb	Mega base (pair)
MCT8	Monocarboxylate transporter 8
Me	Methylation
MMU	Mouse ( <i>Mus musculus</i> ) chromosome
mRNA	messenger RNA
<i>Myh11</i> /Myh11/MYH11	Myosin heavy chain 11
ncRNA	Non-coding RNA
<i>Neb</i> /Neb/NEB	Nebulin
NELF	Negative elongation factor
NKX2.1	NK2 Homeobox 1
nRNA	Nascent RNA
nRNP	Nascent ribonucleoprotein particle
P1	Postnatal day 1
P14	Postnatal day 14
PAX8	Paired box gene 8
PCR	Polymerase chain reaction
Pol II	RNA polymerase II
PRC	Polycomb repressive complex
P-TEFb	Positive transcription elongation factor b
PTM	Posttranslational modification
RNA	Ribonucleic acid
<i>Rpl41</i> /Rpl41/RPL41	60S ribosomal protein L41
RRM	RNA recognition motif
SAM	Synergistic activation mediator
SC35	Splicing component 35 kDa
SE	Super-enhancer
SR	serine and arginine (rich)
Sry	Sex determining region Y
SunTag	SUPerNova tag
T3	Triiodothyronine
T4	Thyroxine

TAD	Topologically associating domain
TF	Transcription factor
<i>Tg</i> / <i>Tg</i> / <i>TG</i>	Thyroglobulin
THRb1	Thyroid hormone receptor beta 1
TL	Transcription loop
TPM	Transcripts per million
TREX	Transcription export complex
TRH	Thyrotropin releasing hormone
TRHR	Thyrotropin releasing hormone receptor
TSH	Thyroid stimulating hormone
TSHR	Thyroid stimulating hormone receptor
TSS	Transcription start site
<i>Ttn</i> / <i>Ttn</i> / <i>TTN</i>	Titin
TTS	Transcription termination site
VPR	VP64-p65-Rta, transcriptional activator