Aus dem Lehrstuhl für Physiologische Chemie im
Biomedizinisches Centrum (BMC)
der Ludwig-Maximilians-Universität München
Vorstand: Prof. Andreas G. Ladurner, PhD

Regulation of Transcription and Stability of Repetitive DNA in S. pombe

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von
Thomas Sander van Emden
aus
Amstelveen, the Netherlands

Jahr
2021

Aus dem Lehrstuhl für Physiologische Chemie im
Biomedizinisches Centrum (BMC)
der Ludwig-Maximilians-Universität München
Vorstand: Prof. Andreas G. Ladurner, PhD

Regulation of Transcription and Stability of Repetitive DNA in S. pombe

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von
Thomas Sander van Emden
aus
Amstelveen, the Netherlands

Jahr
2021

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Betreuer:	Prof. Dr. Andreas G. Ladurner
Zweitgutachter:	PD Dr. Philipp Korber
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	20.05.2022

Table of content

Author's note

A number of figures in this dissertation have been published elsewhere, namely in van Emden et al., (2019) and van Emden & Braun (2019). The respective publication is noted in the legends of the figures to which this applies. Furthermore, the work by co-first author Dr. Marta Forn in the publication van Emden et al., (2019) is described and discussed in this dissertation to allow a consistent flow throughout the presentation of the results. This work is cited as van Emden et al., (2019) and the text will indicate the contribution of co-first author Marta Forn.

Abstract (English):

Repetitive DNA is found in all kingdoms of life and in many species makes up a substantial portion of the genome. Repetitive DNA has many functions, such as: regulatory roles, encoding for long non-coding RNA and it makes up structural components of chromosome. On the other hand, repetitive DNA is unstable and thus a threat to genome stability. Therefore, repetitive DNA needs to be controlled; however, many aspects of this regulation remain elusive.

The genome of the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) has several repetitive regions; among those are the subtelomeres and long terminal repeats (LTRs) of retrotransposons. In the first part of my thesis, I unveiled novel insights into the regulation of the telomere associated sequence (TAS) regions. These repetitive DNA regions are present at most chromosomal ends of *S. pombe*. Previous work in our research group by Dr. Marta Forn has shown that the TAS regions are marked by low nucleosome occupancy. I demonstrated that an ectopic TAS fragment establishes nucleosome levels that are comparable with endogenous TAS regions. Thus, the nucleosome instability is an innate property that possibly originates from the high A-T content of the TAS region.

Furthermore, my results showed that shelterin, a telomere protecting protein complex, counteracts the low levels of nucleosome occupancy in the TAS regions. This is observed both at the endogenous and ectopic TAS regions, demonstrating that shelterin-dependent nucleosome stability is position-independent. Additionally, I showed that the subtelomeric sequences are sufficient to support recombination between subtelomeres in absence of shelterin. Such genome instability may not be desirable for normal cell growth, but provides a mechanism to maintain telomere stability in the absence of shelterin. This suggests that the TAS regions and shelterin dynamically regulate telomere maintenance.

Whereas the TAS regions are located at the subtelomeres, the LTRs are present throughout the genome. Several factors are known to contribute to the repression of LTRs but to a large extent it still is unclear if and how these factors cooperate. In the second part of my dissertation, I described how distinct regulators have different specificity towards subsets of LTRs but that redundancy also exists between the different regulatory pathways.

The inner nuclear membrane protein Lem2 is one of the known regulators of LTRs but the underlying mechanism remains unclear. Lem2 was shown to contribute to heterochromatin silencing and perinuclear localization of centromeres. Notably, I found that Lem2 does not regulate LTRs through heterochromatin or by tethering them to the nuclear periphery. Rather, I found that, at least for a subset of LTRs, Lem2 acts cooperatively with the exosome-targeting factor Red1 involved in RNA degradation.

Finally, to identify novel LTR regulators, I established an LTR-specific reporter assay that allowed me to perform genome-wide genetic screens. Several factors that I identified have an undescribed function in LTRs silencing, unveiling novel potential of regulatory pathways.

Zusammenfassung (Deutsch):

Repetitive DNA findet man in allen Lebensformen und bei vielen Arten macht sie einen wesentlichen Teil des Genoms aus. Repetitive DNA hat viele Funktionen, wie zum Beispiel regulatorische Aufgaben, Kodierung für lange nicht-kodierende RNAs (englisch: long non-coding RNAs, lncRNAs), und sie sind an der Ausbilung der Chromosomstruktur beteiligt. Andererseits ist repetitive DNA instabil und damit eine Bedrohung für die Genomstabilität. Aus diesem Grund wird repetitive DNA kontrolliert, jedoch sind viele Aspekte dieser Regulation bislang wenig verstanden.

Das Genom der Spalthefe *Schizosaccharomyces pombe* (*S. pombe*) weist mehrere repetitive Regionen auf. Darunter fallen die Subtelomere und retrotransposale Elemente, die als LTRs (englisch: long terminal repeats, LTRs) bekannt sind. Im ersten Teil meiner Dissertation gewann ich neue Einblicke in die Regulation der Telomer-assoziierten Sequenzregionen (englisch: telomere associated sequence, TAS). Diese repetitiven DNA-Regionen sind an den meisten chromosomalen Enden von *S. pombe* vorhanden. Frühere Forschungsergebnisse von Dr. Marta Forn in unserer Arbeitsgruppe haben gezeigt, dass die TAS-Regionen durch eine geringe Nukleosomendichte gekennzeichnet sind. Ich konnte zeigen, dass die ektopische Integration eines TAS-Fragment die Chromatinstruktur verändert und Nukleosomendichte vergleichbar ist mit endogenen TAS-Regionen. Somit ist die Nukleosomeninstabilität eine intrinsische Eigenschaft, die möglicherweise auf den hohen A-T-Gehalt der TAS-Region zurückzuführen ist. Darüber hinaus zeigten meine Ergebnisse, dass Shelterin, ein Telomer-protektiver Proteinkomplex, der geringen Nukleosomendichte in den TAS-Regionen entgegenwirkt. Dies wird sowohl an den endogenen als auch an den ektopischen TAS-Regionen beobachtet, was zeigt, dass die Shelterin-abhängige Nukleosomenstabilität unabhängig ist von der chromosomalen Position. Ich konnte außerdem zeigen, dass in Abwesenheit von Shelterin subtelomere DNA-

Sequenzen eine Rekombination zwischen Subtelomeren fördern Diese Genom-Instabilität ist für das normales Zellwachstum vermutlich nicht von Vorteil, könnte aber als Mechanismus dienen, um die Telomerstabilität in Abwesenheit von Shelterin aufrechtzuerhalten. Dies deutet darauf hin, dass die TAS-Regionen und Shelterin zusammen die Erhaltung der Telomere dynamisch regulieren.

Im Gegensatz zu den TAS-Regionen, die sich an den Subtelomeren befinden, sind die LTRs im gesamten Genom vorhanden. Es ist bekannt, dass mehrere Faktoren zur Repression von LTRs beitragen, aber es ist bislang unklar, ob und wie diese Faktoren zusammenwirken. Im zweiten Teil meiner Dissertation konnte ich zeigen, dass bestimmte Regulatoren eine Spezifität für unterschiedliche LTR-Untergruppen aufweisen. Gleichzeitig besteht aber auch Redundanz zwischen den verschiedenen Regulationswegen.

Das innere Kernmembranprotein Lem2 ist an der Regulation von LTRs beteiligt, der zugrunde liegende Mechanismus hierfür ist bis dato aber unbekannt. Von Lem2 ist auch bekannt, dass es der Ausbildung von Heterochormatin beteiligt ist und zur perinukleären-Lokalisierung von Centromeren beiträgt. Ich konnte jedoch zeigen, dass die Funktion von Lem2 in der Regulation von LTRs nicht durch Heterochromatin oder durch Anbinden an die Kernperipherie vermittelt wird. Des weiteren konnte ich zeigen, dass für eine Untergruppe von LTRs Lem2 zusammen mit dem RNA Abbaufaktor Red1 kooperiert.

Um weitere LTR-Regulatoren zu identifizieren, habe ich außerdem einen LTR-spezifischen Reporter-Assay etabliert, der es mir ermöglichte, genomweite genetische Screens durchzuführen. Viele der identifizierte Faktoren haben bisher unbeschriebene Funktionen in der Repression von LTRs und stellen daher ein Potenzial für neue regulatorische Funktionswege dar.

Samenvatting (Nederlands):

In veel soorten vormt repetitief DNA een substantieel deel van het genoom. Repetitief DNA heeft vele functies, zoals: regulerende rollen, coderen voor lang niet-coderend RNA (Engels: long non-coding RNAs, lncRNAs) en het vormt structurele componenten van chromosomen. Aan de andere kant is repetitief DNA ook onstabiel en dus een bedreiging voor de stabiliteit van het genoom. Daarom wordt repetitief DNA gereguleerd; vele aspecten van deze regulatie zijn echter onbekend.

Het genoom van de splijtingsgist *Schizosaccharomyces pombe* (*S. pombe*) heeft verschillende repetitieve regio's, bijvoorbeeld de subtelomeren en de lange terminale herhalingen (Engels: long terminal repeats, LTRs) van retrotransposons. In het eerste deel van mijn proefschrift onthul ik nieuwe inzichten in de regulatie van de telomeer geassocieerde sequentie (Engels: telomere associated sequence, TAS) regio's. Deze zich repeterende DNA regio’s vindt men in de meeste chromosomale uiteinden van *S. pombe*. Eerder onderzoek van Dr. Marta Forn in onze onderzoeksgroep heeft aangetoond dat de TAS regio's worden gekenmerkt door een lage nucleosoombezetting. Ik heb aangetoond dat een ectopisch TAS fragment een nucleosoombezetting heeft die vergelijkbaar is met die van endogene TAS regio's. De nucleosoominstabiliteit is dus een eigenschap van de TAS regio’s en wordt mogelijk veroorzaakt door het hoge A-T-gehalte van de TAS regio’s. Bovendien toonden mijn resultaten aan dat shelterin, een telomeer beschermend eiwitcomplex, de lage nucleosoombezetting in de TAS regio's tegengaat. Ik heb dit zowel in de endogene als in de ectopische TAS regio’s waargenomen. Dit toont aan dat shelterin-afhankelijke nucleosoomstabiliteit positie-onafhankelijk is. Verder toon ik aan dat de DNA sequentie van de subtelomeren, in de afwezigheid van shelterin, voldoende is om recombinatie tussen subtelomeren mogelijk te maken. Een dergelijke instabiliteit van het genoom is niet wenselijk tijdens normale celgroei, maar biedt een mechanisme om de continuïteit van

telomeer te behouden in de afwezigheid van shelterin. Dit suggereert dat de TAS regio's en shelterin de telomeer dynamisch reguleren.

Waar de TAS regio's zich alleen in de subtelomeren bevinden, zijn de LTRs door het hele genoom aanwezig. Het is bekend dat verschillende factoren bijdragen aan de onderdrukking van LTR transcriptie, maar het was onduidelijk of en hoe deze factoren samenwerken. In het tweede deel van mijn proefschrift laat ik zien dat verschillende regulatoren specifieke groepen LTRs reguleren, maar ook dat er redundantie bestaat tussen de verschillende LTR regulerende factoren.

Het binnenste kernmembraaneiwit Lem2 is een van de bekende regulatoren van LTRs, maar het onderliggende mechanisme blijft onduidelijk. Lem2 draagt ook bij aan het reguleren en lokaliseren van heterochromatine. Ik heb echter aangetoond dat Lem2 LTRs niet reguleert via heterochromatine of door ze aan de nucleaire periferie te binden. Daarentegen heb ik ontdekt dat Lem2 eenzelfde subgroep van LTRs reguleert als RNA degradatie factor Red1.

Verder heb ik een LTR-specifieke reportertest opgezet waarmee ik een genoom-brede genetische screening kon uitvoeren om LTR regulatoren te identificeren. Veel van de door mij geïdentificeerde factoren hadden nog geen beschreven functie in het reguleren van LTRs en kunnen daarom duiden op nieuwe mechanisme die LTRs onderdrukken.

List of figures

List of tables

List of abbreviations

5-FOA	5-Fluoroorotic Acid
BLAST	basic local alignment search tool
bp	base pairs
C. elegans	*Caenorhabditis elegans*
ChIP-qPCR	Chromatin immunoprecipitation quantitative polymerase chain reaction
ChIP-seq	Chromatin immunoprecipitation sequencing
CLRC	Clr4-Rik1-Cul4
DNA	deoxyribonucleic acid
DSR	determinant of selective removal
E. coli	*Escherichia coli*
ENCODE	Encyclopedia of DNA Elements
ERV	endogenous retroviruses
H3K9me(2)	(di-)methylation of lysine 9 of histone H3
H3K36me	methylation of lysine 36 of histone H3
HDAC	histone deacetylase
HIRA	Histone Regulator A
HMT	histone methyltransferase
Hoods	heterochromatin domains
HP1	heterochromatin protein 1
kb	kilo base pairs (1000 base pairs)
KRAB-ZFPs	Krüppel-associated box zinc finger proteins
LB	Luria-Bertani
LINE	long interspersed nuclear elements
lncRNA	long non-coding RNAs
LTR	long terminal repeat
MSI	microsatellite instability
ncRNA	non-coding RNA
NGS	next generation sequencing

O/N	overnight
ORC	origin recognition complex
PCR	polymerase chain reaction
PolII-S5P	transcribing polymerase II (phosphorylated at serine 5)
qPCR	quantitative polymerase chain reaction
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RPM	rounds per minute
RT	room temperature
RT-qPCR	reverse transcription quantitative polymerase chain reaction
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SH	subtelomeric homologous
SHREC	Snf2-like/HDAC-containing repressor complex
SINE	short interspersed nuclear elements
siRNA	small interfering RNA
ssDNA	single stranded DNA
TAS	telomere associated sequence
TE	transposable element
TERRA	telomere-repeat encoding RNA
Tf	transposon of fission yeast
TFBS	transcription factor binding sites
wtf	with Tf

1. Introduction

Repetitive DNA is found in every kingdom of life and can make up a substantial portion of an organism's genome. For example, it is estimated that more than two-thirds of the human genome is repetitive (de Koning et al. 2011). Despite its abundance, for many years repetitive DNA has been neglected by researchers because it seemingly did not have any function as it lacks protein-coding sequences (Orgel & Crick 1980). Repetitive DNA was even termed "junk DNA" (Ehret & De Haller 1963). The Encyclopedia of DNA Elements (ENCODE) project is one of the best known examples of studies that aimed to find functional elements in the "junk DNA" (ENCODE Project Consortium et al. 2007). ENCODE scholars defined a functional element as "a discrete genome segment that encodes a defined product (for example, protein or non-coding RNA (ncRNA)) or displays a reproducible biochemical signature (for example, protein binding, or a specific chromatin structure)" (ENCODE Project Consortium 2012). The ENCODE project predicts that about 80% of the human genome sequence meets these criteria, although some researchers believe that the term "functional" is not applied stringently enough, overestimating the amount of functional DNA (Graur et al. 2013). Thus, despite the colossal efforts of projects such as ENCODE, many questions remain with regards to the functions of repetitive DNA.

1.1 Repetitive DNA

Repetitive DNA consists of identical or similar sequences of varying length that are repeated in the genome, and it is classified in two categories: tandem and interspersed repeats. Tandem repeats consist of identical or similar DNA units, repeated from head to tail and may have originated and evolved through slipped strand mispairing, unequal crossover or gene conversion (Levinson & Gutman 1987; Haber & Louis 1998). On the

other hand, interspersed repeats consist of identical or similar sequences spread throughout the genome, and are often remnants of transposable elements (TEs).

1.1.1 Tandem repeats

The group of tandem repeats is composed of satellite DNA, minisatellites and microsatellites. Due to their repetitiveness, tandem repeats have a base composition that is distinct from bulk DNA, which results in the formation of a "satellite" band when genomic DNA is separated on a density gradient, explaining the names of these repeats (Kit 1961).

1.1.1.1 Satellite DNA

Satellite DNA consists of very large arrays of tandem repeats; found for example in centromeres and telomeres. The centromeres are unique chromosomal regions that mark the sites for kinetochore assembly during cell division (Schalch & Steiner 2017). The kinetochore forms a platform for spindle microtubule attachment, which is necessary for faithful separation of the two sister chromatids. Kinetochore or spindle assembly defects can lead to aneuploidy and promote tumorigenesis (Weaver et al. 2007). Conversely the telomeres are found on the extremities of chromosomes. Human telomeres consist of 10-15 kilo base pairs (kb) of TTAGGG repeats and are protected by a protein complex called shelterin (Pfeiffer & Lingner 2013; Webb & Zakian 2012; Heidenreich & Kumar 2017). Components of the shelterin complex recruit telomerase, which in turn prevents the shortening of the telomeres that is inherent to DNA replication (Xin et al. 2007; Miyoshi et al. 2008; Abreu et al. 2010). Furthermore, shelterin protects telomeres from being recognized as double-strand breaks, preventing linear chromosomal fusions (Palm & de Lange 2008).

1.1.1.2 Minisatellites and microsatellites

Besides centromeres and telomeres, other tandem repeats can be found dispersed all over the genome. Based on their repeat length, they are classified as minisatellites or microsatellites.

Minisatellites are tandem repeats in which the repeat unit ranges between 10 and 50 base pairs (bp). Minisatellites pioneered the use of DNA in forensic science, although later, with more developed technology, the use of microsatellites became more practical (Jeffreys, Wilson & Thein 1985a; Jeffreys, Wilson & Thein 1985b; Panneerchelvam & Norazmi 2003).

Microsatellites are tracts of tandem repeats in which the repeat unit ranges between 2 and 5 bp. Typically, these repeat units are repeated 5-50 times (Richard et al. 2008). They mutate at a frequency more than 10 orders of magnitude greater than regular point mutations (Verstrepen et al. 2005). This is because microsatellites are prone to expansion and contraction; as a consequence, the number of repetitive units found in tandem is constantly changing. This makes the study of microsatellites interesting for many different fields of science, for example for forensic science to study kinship and genealogy (Lászik et al. 2000; Kopelman et al. 2009; Kayser 2017). However, with the advent of cost-effective, high throughput genotyping and sequencing platforms, the role of microsatellites in forensic sciences is diminishing. Nevertheless, microsatellites remain an active topic of research, as microsatellite instability (MSI) is a hallmark of certain cancer types (Ghidini et al. 2020). For example, the Lynch syndrome is a genetic condition characterized by inherited mutation in DNA mismatch repair genes (Lynch et al. 2009). As a result, individuals with this mutation have increased MSI and a high risk of developing cancer. This type of cancer is diagnosed through the analysis of the size of microsatellites in healthy and tumor tissues (Vaksman & Garner 2015).

Besides being a biomarker for cancer, expansion of microsatellites itself can be a causative agent of disease. More than 40 neurological, neurodegenerative and neuromuscular disorders are linked to repeat instability. The majority of these disorders are caused by instability of shorter repeats (tri, tetra and pentanucleotide repeats)(Pearson et al. 2005). A trinucleotide repeat expansion in the huntingtin protein-coding gene leads to a protein variant causing brain cell damage (The Huntingtons Disease Collaborative Research Group 1993). Patients with Huntington's disease experience symptoms of neural degeneration that start with subtle problems in mental abilities and then progress to dementia (Frank 2014; Dayalu & Albin 2015).

1.1.2 Interspersed repeats

1.1.2.1 Interspersed repeat classification

Besides tandem repeats, the human genome is made up of 45% interspersed repeats of TE origin (Lander et al. 2001). Remnants of TEs also significantly contribute to the genome content of other mammalian genomes, for example the genomes mice and dogs (37.5% and 41%, respectively), but TEs are also highly abundant in non-mammalian genomes, such as the genomes of *Zea mays* (maize) and *C. elegans* (66% and 12%, respectively)(Mouse Genome Sequencing Consortium et al. 2002; Lindblad-Toh et al. 2005; Haberer et al. 2005; C. elegans Sequencing Consortium 1998).

Different properties of TEs can be used to classify them, for example phylogeny or transposition requirements, but the first and most common classification is based on the nature of the transposition intermediate: RNA or DNA (Finnegan 1989; Seberg & Petersen 2009). Over the years, this system was further refined by adding subclasses, orders and families (Wicker et al. 2007). In this broadly accepted classification, TEs belonging to class I require retrotranscription in order to transpose and are referred to as retrotransposons. Class I TEs include, amongst others, LTR retrotransposons, long interspersed nuclear

elements (LINE) and short interspersed nuclear elements (SINE). LTR retrotransposons and endogenous retroviruses (ERVs) have the same evolutionary origin and thus share many similarities. Both are flanked by LTRs that form the promoter and terminator of this type of retroelement. While the exact content of the retroelement may differ from species to species, the core components are the *gag*, *pro*, *pol* and *env* genes (Stoye 2012). The *gag* gene encodes the core structural component of the immature viral-like particle. The gene product of the *pro* gene contributes to maturation of this particle. The *pol* gene codes for a reverse transcriptase and all other elements required for retrotranscription. The fourth gene, *env*, often separates LTR retrotransposons from ERVs. The *env* gene encodes for the glycoproteins that is part of the lipid bilayer of the retrovirus, and is required for viral infections (Coffin et al. 1997). Through deletion or mutation of the *env* gene, ERVs can lose their ability to form viral particles and become LTR retrotransposons (Katoh & Kurata 2013). As mentioned above, class I TEs require retrotranscription for their activity, following a "copy and paste" mechanism (Sotero-Caio et al. 2017). By copying their genomic material class I TEs can, very efficiently, spread throughout eukaryotic genomes (Wells & Feschotte 2020).

Class II TEs are referred to as DNA transposons. They employ a less efficient mechanism of spreading known as "cut and paste". Instead of making a copy of their viral genome, they excise it directly from the host's genome and insert it elsewhere. The lower efficiency of spreading is reflected in the abundance of class I and class II elements: 42% of the human genome is composed of class I, whereas 3% is of class II origin (Lander et al. 2001).

1.1.2.2 Consequence of TE integration

Once a TE has integrated in the genome of a host, both host and TE have to adapt to the new situation. Recent studies have shown that host and TE can co-evolve and that TEs can take on regulatory roles in the host genome (Ågren & Wright 2011; Cosby et al. 2019; Choudhary et al. 2020). This is not surprising, as LTRs, in their native function, are important for TE expression. Furthermore LTRs often contain transcription factor binding sites (TFBS), which are recognized by key regulators such as p53, OCT4, SOX2 and NANOG (Wang et al. 2007; Kunarso et al. 2010). Through these TFBS LTRs can regulate gene expression of the host genome (Emera & Wagner 2012; Sundaram et al. 2014). Additionally, recent studies revealed, for particular TEs, additional important functions, such as the generation of long non-coding RNAs (lncRNA) and contribution to the 3D chromosomal structure (Kapusta et al. 2013; Diehl et al. 2020). Moreover, TEs not only take on regulatory roles but may also serve as a reservoir for diversity, which is important during evolution (Fondon & Garner 2004).

Like tandem repeats, interspersed repeats are also implicated in pathogenesis. For example, misregulation of specific TEs has been detected in neurodegenerative diseases (W. Li et al. 2012). Furthermore, TEs form a threat to the integrity of the host genome. For instance, insertion of retrotransposons can be mutagenic to protein-coding genes, and gene regulation can be disturbed by LTRs acting as alternative promoters or enhancers (Girard & Freeling 1999). Finally, sequence similarity between two LTRs in close proximity to each other allows for homologous recombination and, thus, can lead to chromosomal rearrangements (Garfinkel et al. 2005).

1.1.3 The role of Heterochromatin in TE silencing

Repetitive DNA is often found in specific chromatin domains that are highly compact and mostly transcriptionally inactive, which is believed to protect the genome from harmful

effects. These compact chromatin domains are called heterochromatin, which is the opposite of euchromatin, the open and transcribed form of chromatin. Both forms of chromatin are characterized by modifications of specific residues on specific histone tails. Prime examples of heterochromatin marks are methylation of histone H3 at lysine 9 and lysine 27 (H3K9me and H3K27me). Several core components of the heterochromatin machinery contribute to the establishment, spreading and maintenance of heterochromatin (Allshire & Madhani 2018). The methylation of lysine 9 on the histone H3 tail is established by histone methyltransferases (HMTs). The human SUV39H1 was the first described enzyme to specifically catalyze this reaction through its SET domain (Rea et al. 2000). Since then, more HMTs have been identified in various organisms. Methylated H3K9 is recognized by and bound by so called readers (Yun et al. 2011). The SUV39H1 homologs combine reading and writing as they contain both a Chromo domain that recognizes H3K9me and a catalytically active Set domain. Additionally, members of the heterochromatin protein 1 (HP1) family bind H3K9me and recruit effector proteins such as histone deacetylases (HDACs) that remove the active acetylation marks (James & Elgin 1986; Bannister et al. 2001; Lachner et al. 2001; C. L. Zhang et al. 2002).

At the telomeres, transcription of telomere-repeat encoding RNA (TERRA), a lncRNA encoded in the telomeres of humans and yeast, is required for the establishment of heterochromatin (Azzalin et al. 2007; Luke et al. 2008; Schoeftner & Blasco 2008). Upon transcription, TERRA is recruited to telomeres through interaction with the core shelterin proteins. Notably, TERRA depletion leads to a loss of H3K9me at telomeres and decreased recruitment of the Origin Recognition Complex (ORC)(Deng et al. 2009). The ORC, primary known for its function in initiation of DNA replication, has been shown to be important for heterochromatin-related functions at centromeres but its role at telomeres remains elusive (Prasanth et al. 2004; Shimada & Gasser 2007).

To silence TEs, different organisms have developed different methods. In the germline of *C. elegans*, transposons are silenced by RNAi (Sijen & Plasterk 2003). In many other organisms H3K9me-mediated heterochromatin is involved in the silencing of TEs (Slotkin & Martienssen 2007). For example, in mouse embryonic stem cells, mutations in the HMT encoding gene *Suv39,* lead to upregulation of TEs (Martens et al. 2005). In human cells, several factors have been shown to silence TEs, among which are the Krüppel-associated box zinc finger proteins (KRAB-ZFPs). Despite extensive ChIP-seq studies showing that ~2/3 of human KRAB-ZFPs bind TEs, the function of most KRAB-ZFPs remains largely elusive (G. Wolf et al. 2015; Schmitges et al. 2016). Only the KRAB-ZFP protein ZFP809 was directly shown to form a stable complex with TRIM28 (Tripartite Motif Containing 28), a regulator of transcription, in order to repress TE activity. The TRIM28 complex recruits heterochromatin factors such as HP1, SETDB1 (a HMT) and nucleosome remodeling and deacetylase (NuRD)(Sripathy et al. 2006; D. Wolf & Goff 2007; D. Wolf et al. 2008). DNA methylation is another extensively studied mechanism that is used by many eukaryotes to target the heterochromatin machinery to the TEs. During the first days of mammalian embryogenesis, ERV DNA is *de novo* methylated (Rowe & Trono 2011). Interestingly, these early waves of DNA methylation seem to require a sequence-specific binding of KRAB-ZFP to the ERVs and recruitment of TRIM28 and SETDB1 (Rowe et al. 2013).

1.1.4 Challenges in studying repetitive DNA

Repetitive DNA remains an active field of research, as many facets, such as its regulation and function, are not fully understood. Our limited knowledge of the function of repetitive DNA is partially caused by the difficulty of studying repetitive sequences. For example, *de novo* genome assembly that relies on first (Sanger) or next generation sequencing

(NGS) technology (454/SBS and Illumina) is hampered by short read length; resulting in the failure to capture the full extent of the repetitive DNA sequence. With the advent of long read sequencing technologies, some of these challenges can be overcome (Shahid & Slotkin 2020). Third generation (Oxford Nanopore and PacBio) sequencing technologies provide longer read length but are more error prone and more expensive. However, nowadays both cost and error rates have dropped and third generation sequencing technologies are successfully used to resolve gaps in genome assemblies that could not be resolved by the previous sequencing technologies (Müller et al. 2018). But even with a fully assembled genome studying repetitive DNA remains challenging. Methods that rely on mapping NGS reads to genomes or transcriptomes, such as ChIP-seq or RNA-seq, are hampered by the limited ability to determine the exact binding sites of the immunoprecipitated protein or origin of the transcript, respectively. Moreover, also more traditional methods such as expression analysis by reverse transcription quantitative polymerase chain reaction (RT-qPCR)(primer design for specific loci) and reporter gene insertion (exact location of insertion) struggle to provide insights into repetitive DNA.

1.2 Repetitive DNA of *S. pombe* subtelomeres

1.2.1 Studying repetitive DNA in *S. pombe*

The single cellular eukaryote fission yeast *S. pombe* is an ideal genetic model system to study repetitive DNA, as it contains both tandem and interspersed repeats. Furthermore, many of the structural components and enzymes involved in mammalian heterochromatin are conserved in *S. pombe*.

Tandem repeats regions in the genome sequence of *S. pombe* are found, amongst others, in the telomeres. The origin and integration of tandem repeats at the chromosome ends remain a matter of speculation (de Lange 2004; Nosek et al. 2006; Peska & S. Garcia

2020). Compared to the human telomeres (10 to 15 kb), the telomeres of *S. pombe* are relatively short in length, with only 300 bp of the repeat unit $G_{2\text{-}6}$TTAC[A] (Sugawara 1988; Canela et al. 2007; Dehé & Cooper 2010). Like their human counterparts, the telomeres of *S. pombe* are bound by the shelterin complex.

1.2.2 The shelterin complex and its binding partners in *S. pombe*

In *S. pombe*, the shelterin complex consists of six proteins. Pot1 is a ssDNA (single stranded DNA) -binding protein that binds the 3' telomeric overhang, connecting the outermost portion of the telomere with shelterin. Pot1 further interacts with the shelterin subunit Tpz1 (Miyoshi et al. 2008). Another member of the shelterin complex is Taz1, which binds double stranded DNA and recruits shelterin subunit Rap1 (Cooper et al. 1997). The fifth shelterin protein, Poz1, bridges the Pot1-Tpz1 and Taz1-Rap1 subcomplexes (Takai et al. 2011). Unlike human shelterin, the *S. pombe* shelterin complex contains a sixth protein called Ccq1, which interacts with Tpz1 (Harland et al. 2014). Ccq1 recruits the Snf2-like/HDAC-containing repressor complex (SHREC), a homolog of the mammalian NuRD complex, to the telomere (Sugiyama et al. 2007). SHREC is composed of the ATP-dependent nucleosome remodeler Mit1, the HDAC Clr3, and the methyl-CpG binding domain-like protein Clr2, which are assembled together on the scaffold protein Clr1 (Sugiyama et al. 2007; Job et al. 2016). Besides its role in telomeric regulation, shelterin promotes silencing of reporter genes inserted adjacent to the telomere (Cooper et al. 1997; Kanoh et al. 2005; Moser et al. 2015).

1.2.3 Organization of *S. pombe* subtelomeres

In *S. pombe*, subtelomeres are a 100 kb region directly adjacent to the telomeres. They are partially repetitive and share homology between the different chromosomal arms, making it difficult to determine their individual DNA sequence. The subtelomeres are split into

two ~50 kb regions, known as the knob region and the subtelomeric homologous (SH) region (Tashiro et al. 2017; Sugawara 1988). The knob regions are located telomere-distal, between ~50-100 kb on both arms of chromosome I and II. These highly dense chromatin region was originally discovered by super-resolution fluorescence microscopy as the most dense chromatin regions of the *S. pombe* genome, hence its name (Matsuda et al. 2015). The SH regions are located between the knob regions and the telomeres. In contrast to the knob regions, the SH regions of different chromosome arms are homologous, yet not identical (Chaudari & Huberman 2012; Oizumi et al. 2021). Notably, removal of all SH regions demonstrated that these sequences are not critical for mitotic, and meiotic growth or telomere length control (Tashiro et al. 2017). However, the SH regions do play an important role in *S. pombe* survival upon telomere shortening due to disruption of *trp1*$^{+}$, the gene encoding telomerase. When faced with shortened telomeres, *S. pombe* has three mechanisms of survival. The most common survival strategy is self-circularization of chromosomes (Nakamura et al. 1997). A less favored mechanism is through homologous recombination between chromosomal arms, which depends on DNA recombination protein, Rad52 (Nakamura et al. 1998). Finally, non-telomeric heterochromatic regions can be used to replace the telomeres in a mechanism called heterochromatin amplification-mediated and telomerase-independent (Jain et al. 2010).

The TAS refers to the most telomere proximal 10 kb of the SH region (Sugawara 1988). Homologous TAS regions have been identified in many organisms, such as plants, birds and budding yeast (Vrbsky et al. 2010; Solovei et al. 1994; Luke et al. 2008). In *S. pombe*, the TAS regions are present on at least four of the six chromosomal arms of laboratory strains, however, strains isolated from nature contain TAS regions on 5 chromosomal arms (Oizumi et al. 2021). The TAS regions are flanked by the telomeres and telomere-distal by the *tlh*$^{+}$ genes (i.e. *tlh1*$^{+}$, *tlh2*$^{+}$, *tlh3*$^{+}$ and *tlh4*$^{+}$, one on each arm)(Oizumi et al.

2021). Whereas most parts of the SH regions are found in the published genome, large parts of the TAS regions are omitted (Wood et al. 2002). However, early efforts to determine the sequence of the TAS regions have resulted in a series of plasmids that contain parts of the TAS regions. Based on restriction digestion patterns the TAS regions have been spilt up in three regions: TAS1, TAS2 and TAS3. Although the TAS regions are presumed to be repetitive, their precise repeat units are not clearly defined (Sugawara 1988).

The TAS region gives rise to various RNA species (Greenwood & Cooper 2012). One of those is the long non coding TERRA, which at its discovery was a surprise as the telomere and TAS were thought to be transcriptionally silent regions (Azzalin et al. 2007).

1.2.4 Transcriptional regulation of *S. pombe* subtelomeres

Despite their close proximity the silencing of knob and TAS regions is regulated differently. Interestingly, genes such as *clr3*$^+$, *dcr1*$^+$, *clr4*$^+$ and *swi6*$^+$, which encode classic silencing proteins, are not required of knob formation (Grewal 2000; Bühler & Gasser 2009). Instead, methylation of lysine 36 of histone H3 (H3K36me), a histone modification normally associated with the coding region of genes and thus with transcription elongation and actively transcribed regions, is required for the formation and gene silencing of the knob region. In line with this, a point mutant that mimics acetylation of H3K36 disrupts the structure of the knob regions (Bannister & Kouzarides 2011; Martin & Y. Zhang 2005; Kizer et al. 2005; Morris et al. 2005). Furthermore a protein called Sgo2 is important for knob formation and is recruited to the subtelomere by the action of the HMT of H3K36, Set2 (Morris et al. 2005; Tashiro et al. 2016). Deletion of both *set2*$^+$ and *sgo2*$^+$, leads to derepression of genes not only in the knob region but also in the more telomere adjacent SH region (Matsuda et al. 2015; Tashiro et al. 2016). It remains how-

ever unknown how transcription of the TAS region, that codes for the lncRNA TERRA, is regulated.

1.3 Interspersed repeats in *S. pombe*

Besides the tandem repeats at telomeres and subtelomeres, the genome of *S. pombe* also contains interspersed repeats (Wood et al. 2002). These repeats are transposable elements or derived from them. Although retroelements have been very successful in spreading into eukaryotic genomes, they only make up 1.1% of the genome of *S. pombe* and they are all derived from LTR retrotransposons (Bowen et al. 2003).

1.3.1 Organization of LTR retrotransposons and solo LTRs in *S. pombe*

The transposons of fission yeast (Tf) were first described while studying the genomes of various *S. pombe* strains with probes made of unidentified repetitive *S. pombe* DNA. DNA sequences that hybridized to these probes were cloned and sequenced and shown to contain two types of Tfs (Tf1 and Tf2)(Levin et al. 1990). Sequence comparison revealed that the Tfs of *S. pombe* are similar to the Ty3/Gypsy retrotransposon family, which has been extensively studied in *S. cerevisiae* (*Saccharomyces cerevisiae*)(Levin et al. 1990). Notably, analyses of the sequences of the Tf1 and Tf2 LTRs revealed potential TATA boxes that are similar to the highly expressed *adh1*$^+$ gene, suggesting that the LTRs play a role in transcription of the retroelement. Indeed, Tf1 and Tf2 were both found to be transcribed (Levin et al. 1990). More than a decade later, the complete genome sequence of the most common *S. pombe* laboratory strain was published, this strain is different from the strains used to identify Tf1 and Tf2 (Wood et al. 2002). Searching for the Tf sequences in the fully assembled genome sequence revealed that the most common labor-

atory strain contains 13 Tf2 full-length retrotransposons, but no Tf1 full-length retrotransposons (Bowen et al. 2003).

Eukaryotic cells eliminate retrotransposons by homologous recombination between the LTRs of the retrotransposon, leaving solo LTRs in the genome (Shirasu et al. 2000). Indeed, there are 28 and 35 solo LTRs originating from Tf1 and Tf2 retrotransposons, 111 LTR-related sequences, and 75 LTR-related fragments identified in the genome sequence of *S. pombe* (Bowen et al. 2003; Belshaw et al. 2007). Full-length retrotransposons have preferentially integrated in promoter regions of genes (Bowen et al. 2003). This preference is probably a consequence of the more open chromatin structure of promoter regions (Bowen et al. 2003). Conversely, insertion in a gene body is likely detrimental and may have been selected against. To date, 13 full-length retrotransposons and 214 solo LTRs are annotated for the *S. pombe* genome available on the genome browser *PomBase* (Wood et al. 2002; Lock et al. 2019). This annotation is used throughout this work as many other resources, such as annotation files and the *S. pombe* genome browser, are based on the *PomBase* annotation.

1.3.2 Regulation of *S. pombe* LTRs

In *S. pombe*, full-length retrotransposons are transcriptionally repressed by proteins known as CENP-B homologs. This conserved protein family plays a prominent role at centromeres but has also non-centromeric functions. CENP-B homologs mediate the formation of nuclear Tf bodies into which retrotransposons cluster together. Upon deletion of *abp1*$^{+}$, Tf bodies de-cluster, illustrating a critical role in nuclear organization of retrotransposons and CENP-Bs (Cam et al. 2008). The CENP-B protein Abp1 (also referred to as Cbp1) binds directly to LTRs and mediates recruitment of the HDACs Clr3 and Clr6. However, these HDAC mutants show only partial silencing defects compared to

*abp1*Δ, thus recruitment of HDACs does not fully explain transcriptional silencing and recombinatorial repression of retrotransposons by Abp1 (Cam et al. 2008). Interestingly, a *wtf* (with TF) element and several LTR neighboring genes are shown to also be repressed by Abp1 (Cam et al. 2008). Furthermore, *de novo* integration of a Tf1 full-length retrotransposon also leads to upregulation of neighboring genes, but only in the absence of $abp1^{+}$ (Feng et al. 2013). This suggests that Abp1 limits the ability of retroelements to regulate their neighboring genes or that the neighboring genes are regulated by Abp1 as a side effect of retroelement regulation.

Genome-wide analysis of Abp1 binding data revealed a 10 bp A-T rich motif that is important for Tf2 element binding by Abp1. Deletion of this domain reduces Abp1 binding to Tf2 LTRs *in vitro* (Lorenz et al. 2012).

Besides LTR silencing and centromere functions, CENP-B proteins also play a role in DNA replication (Zaratiegui et al. 2011). In *S. cerevisiae*, LTRs and various other genomic features have been marked as fragile or hard-to-replicate sites. These regions are characterized by replication fork stalling and often marked by histone H2A phosphorylation (Szilard et al. 2010). In *S. pombe*, replication fork stalling depends Sap1, which is recruited to LTRs (Mejía-Ramírez et al. 2005; C. Noguchi & E. Noguchi 2007). CENP-B proteins counteract Sap1 and promote replication fork stability, thereby ensuring replication of the LTRs (Zaratiegui et al. 2011).

The H3K4 methyltransferase Set1 was found to play a partially overlapping role with Abp1 in the silencing of LTRs (Lorenz et al. 2012). However, the HMT activity of Set1 does not play a role in regulation of LTRs, since methylation-deficient H3K4 mutants do not exhibit higher LTR expression. Furthermore, the formation of Tf-bodies is independent of Set1 (Lorenz et al. 2012). Thus, how Set1 regulates expression of Tf2 LTRs remains elusive.

Besides the CENP-B-dependent pathway, several other mechanisms contribute to LTR silencing in *S. pombe*. One such mechanism involves Histone Regulator A (HIRA), a histone chaperone complex that deposits nucleosomes preventing the transcription of many genes (H. E. Anderson et al. 2009). Deletion of genes encoding the HIRA components Hip1, Slm9 and the HIRA interacting protein Hip3 causes increased expression of many genes, including the 13 full-length retrotransposons. Additionally, non-coding transcripts from solo LTRs are increased in *hip1*Δ and *slm9*Δ cells (H. E. Anderson et al. 2009; Greenall et al. 2006). *hip1*$^{+}$ and *abp1*$^{+}$, have a synthetic genetic interaction with respect to Tf2 element expression indicating that CENP-B and HIRA proteins regulate Tf2 elements through distinct pathways (Cam et al. 2008).

Additionally, genome-wide binding studies showed that the chromatin remodeler Fft3, which belongs to the Snf2 Fun30/SMARCAD1 subfamily, binds to LTRs. Micrococcal nuclease-based mapping revealed decreased nucleosome occupancy at LTRs in *fft3Δ* cells, suggesting that Fft3 maintains nucleosomes at LTRs (Steglich et al. 2015). By regulating nucleosome occupancy, these Fun30 proteins regulate the transcription start site (TSS) of the Tf2 retrotransposon. Intriguingly, control of nucleosome maintenance at TSS is relieved under stress, allowing for transcription of retrotransposons (Persson et al. 2016). Furthermore, the (sub)nuclear localization of LTRs seems to partially depend on Fft3, as deletion of *fft3*$^{+}$ results in a minor but specific decrease in interaction of LTRs with nuclear membrane proteins (Steglich et al. 2015). The *S. pombe* genome encodes for two other members of this family, Fft1 and Fft2. Fft2 has been shown to function redundantly with Fft3 in the regulation of Tf2 retrotransposon transcription.

The Tf-bodies, that are mediated by Abp1, de-cluster upon deletion of Fun30 family members, it is however unclear if this is regulated through the Abp1 dependent pathway

or represents a district pathway (Persson et al. 2016). Likewise it is unclear if Fun30 family members and CENP-B proteins cooperate in the regulation of LTRs.

The previously discussed silencing pathways repress full-length retrotransposons at the level of transcription. However, retroelements are also regulated post-transcriptionally by the nuclear exosome. This highly conserved protein complex degrades unstable and aberrant transcripts through its 3'→5' exonuclease activity. One of the catalytic subunits of the nuclear exosome is Rrp6. Upon deletion of *rrp6*$^{+}$, several Tf2 retrotransposons produce siRNAs and accumulate H3K9me, forming heterochromatin domains (HOODs)(Yamanaka et al. 2013). Furthermore, mutants of the exosome targeting complex Mtl1-Red1 core (MTREC) trigger the upregulation of transcripts originating from Tf2 solo LTRs and Tf2 retrotransposons (Sugiyama & Sugioka-Sugiyama 2011; N. N. Lee et al. 2013). Red1 does not localize with Abp1 in the Tf-bodies suggesting that Red1 acts independently of Abp1 (Sugiyama & Sugioka-Sugiyama 2011). Contrarily, another study that investigated the transcript level of Tf2 retrotransposons in mutants of *abp1*$^{+}$ and genes encoding exosome targeting factors (i.e. Pab2, Rrp6) found that deletion of the exosome components suppresses the phenotype of the *abp1*$^{+}$ mutants (Mallet et al. 2017). Despite the growing body of knowledge about full-length retrotransposon silencing in *S. pombe*, the mechanism by which solo LTRs are regulated remains largely unknown.

Interestingly, a recent study of the Braun lab identified Lem2 as a regulator of subtelomeric LTR silencing (Barrales et al. 2016). Lem2 is an inner nuclear membrane protein, with 2 domains facing nuclear lumen connected through two transmembrane domains. One of the nucleoplasmic domains of Lem2 contains a LAP2–Emerin–MAN1 (LEM) domain, a highly conserved 40 amino acid helix-extension-helix motif (Brachner & Foisner 2011). Such LEM domains are found more often in lamin-associated proteins, it should be noted however, that unicellular organisms like *S. pombe* do not have a nuclear

lamina (Brachner & Foisner 2011). The N-terminus of Lem2, containing the LEM domain, binds to centromeres, contributing to their tethering to the nuclear periphery (Barrales et al. 2016; Hirano et al. 2018). The C-terminus of Lem2 contains an MAN1–Src1 C-terminal (MSC) domain that contains a winged-helix fold motif. The MSC domain does not interact with centromeres, nevertheless it mediates silencing of pericentromeres and other heterochromatin domains (Barrales et al. 2016). Deletion of *lem2*$^{+}$ leads to decreased association of SHREC with heterochromatin chromatin, hence it has been proposed that Lem2 regulates the silencing of heterochromatin through the interaction with SHREC (Barrales et al. 2016). Furthermore, Lem2 plays a role in degradation of exosome targets including ncRNAs and meiotic transcripts (Caballero et al. 2021). This mechanism is thought to promote the recruitment of nuclear exosome targeting factors to the nuclear periphery. The MSC domain again mediates this function. Besides the recruitment and interaction with repressor factors, Lem2 also contributes to nuclear structure and integrity, for instance by limiting the flow between the nuclear membrane and ER network connected through the outer nuclear membrane. This role of Lem2 has been proposed to control the size of the nucleus, but the exact mechanism remains elusive (Kume et al. 2019).

Other eukaryotes employ different mechanisms to silence LTRs, the best studied examples are DNA methylation and the recruitment of zinc finger proteins. However, homologs to these zinc finger proteins and DNA methylation are not found in *S. pombe.* Other examples of mechanism used by eukaryotes to silence TEs are RNAi and H3K9me. These have been extensively studied in *S. pombe* in the context of heterochromatin, but they seem to play no or only a minor role in LTR silencing (Allshire & Ekwall 2015; Martienssen & Moazed 2015). In particular, genome-wide studies found no enrichment of H3K9me at or near to LTRs and Tf2 retrotransposons and deletion of RNAi components

does not result in an increased expression of neither Tf2 nor LTR neighboring genes (Cam et al. 2005; Hansen et al. 2005).

In conclusion, several factors are known to regulate full-length retrotransposons in *S. pombe* but their role in silencing of LTRs remains poorly understood. Furthermore, no systematic analysis has been performed to identify pathways among the known regulators. Lem2 regulates solo LTRs but it is not known if this is restricted to the subtelomeric LTRs or rather a phenomenon seen for LTRs genome-wide. Additionally, the mechanism by which Lem2 regulates LTRs remains elusive.

1.4 Aim of the present study

Repetitive DNA makes up a significant portion of our genome, yet many facets of its regulation remain elusive. This dissertation explores the regulation of two types of repetitive regions: telomeres and LTRs. Part of the studies presented here continues on the work of Dr. Marta Forn, providing an in-depth analysis of the cause and consequences of low nucleosome density of the TAS regions in *S. pombe*. Beyond, this study explores the role of Ccq1 and the TAS regions in controlling the genomic stability of the subtelomeres.

In the second part, the work described in this dissertation seeks to identify functional pathways among known regulators of LTR expression. Finally, through a hypothesis driven approach and an unbiased genome-wide approach, this study aims to elucidate the mechanism by which inner nuclear membrane protein Lem2 regulates the silencing of LTRs.

2. Materials and Methods

2.1 Bacterial culturing, transformation and storage

2.1.1 Liquid media and media plates for bacterial culturing

Table 1 Luria-Bertani (LB) liquid media

Ingredient	Final concentration
Yeast extract	5 g/L
Tryptone	10 g/L
NaCl	10 g/L

All ingredients are mixed and autoclaved at 121 °C for 20 minutes.

Table 2 LB + Amp liquid media

Ingredient	Final concentration
LB liquid media (Table 1)	-
Ampicillin	100 μg/mL

Ampicillin is added directly to the LB liquid media as prepared in Table 1.

Table 3 LB + Amp plates

Ingredient	Final concentration
Yeast extract	5 g/L
Tryptone	10 g/L
NaCl	10 g/L
Agar (Serva)	1.5%
Ampicillin	100 μg/mL

All ingredients, except ampicillin, are mixed and autoclaved at 121 °C for 20 minutes. Ampicillin is mixed in after autoclaving. 35 mL is used to fill a petri dish (Ø 8.5 cm) for standard usage.

2.1.2 Bacterial culturing

Escherichia coli (E.coli) was grown in standard laboratory conditions, i.e. at 37 °C and in LB media. For detailed media ingredients see Table 1, Table 2 and Table 3. When *E. coli* was grown in test tubes (2 mL), the liquid culture was rotated at 150 rounds per minute (RPM). Generally, *E. coli* was streaked from long-term storage on a LB plate and left to grow at standard conditions for 1 day before liquid cultures we started from single colonies.

2.1.3 Bacterial transformation and plasmid handling

To transform bacteria, competent cells (XL1 Blue (Stratagene) or DH5α (Thermo Fisher Scientific)) were thawed on ice and incubated with the plasmid that the bacteria were to be transformed with. After 15 minutes the bacteria were submitted to a heat shocked for 1 minute at 42 °C and directly transferred to ice. Bacteria were resuspended in 1 mL LB media and allowed to recover at 37 °C for 1 hour while shaking at 500 RPM. Bacteria were then plated at different dilutions on LB + Amp plates and grown overnight (O/N). Single colonies were picked and grown O/N in 2 mL liquid LB + Amp media. Cultures were either mixed with 50% glycerol in a 1:1 ratio for long-term storage at -80 °C or the plasmid was extracted using a Miniprep kit (Metabion, mi-PMN250) according to manufacturer's instruction. Plasmid was either stored at -20 °C or checked by enzymatic digestion or Sanger sequencing (Genewiz).

2.1.4 Plasmids used in this dissertation

Table 4 Plasmids used in this dissertation

Plasmid #	Plasmid name	genotype	source
ESB098	pFA6a-kanMX6	ori, ampR, kanR	(Bähler et al. 1998)
ESB096	pFA6a-natMX6	ori, ampR, natR	(Hentges et al. 2005)
ESB388	pFA6a-hygMX6	ori, ampR, hygR	(Hentges et al. 2005)
ESB558	pLSB		(Torres-Garcia et al. 2020)
ESB570	pLSB-sgRNA-ura4-full deletion		This study
ESB508	pJK148	ori, ampR, leu1	(Keeney & Boeke 1994)
ESB510	pJK148 TAS1 (800bp)	ori, ampR, leu1, TAS1 (800bp)	(van Emden et al. 2019)
ESB569	pBluescript ii sk+	ori, ampR	(Alting-Mees & Short 1989)
ESB624	pBluescript ii sk+ 8xLTR reporter	ori, ampR 8xLTR reporter	This study

2.2 *S. pombe* culturing and storage methods

2.2.1 Liquid media and plates for *S. pombe* growth

Table 5 Edinburgh minimal media (EMM) –Leu liquid media

Ingredient	Final concentration
EMM-GLuc	12.3 g/L
SP supplements -Leu -Ura	0.6 g/L
Uracil	0.2 g/L
Glucose	2%

All ingredients, except glucose, are mixed and autoclaved at 121 °C for 20 minutes. Glucose is mixed in after autoclaving.

Table 6 EMM –Leu plates

Ingredient	Final concentration
EMM-GLuc	12.3 g/L
SP supplements -Leu -Ura	0.6 g/L
Uracil	0.2 g/L
Agar (Serva)	2%
Glucose	2%

All ingredients, except glucose, are mixed and autoclaved at 121 °C for 20 minutes. Glucose is mixed in after autoclaving. 35 mL is used to fill a petri dish (Ø 8.5 cm) for standard usage.

Table 7 Yeast extract with supplements (YES) liquid media

Compound	Final concentration
Yeast extract	5 g/L
SP supplements	1 g/L
KH_2PO_4	~56 mM (used to adjust pH to 5.7)
Glucose	3%

All ingredients, except glucose, are mixed and autoclaved at 121 °C for 20 minutes. Glucose is mixed in after autoclaving.

Table 8 YES plates

Compound	Final concentration
Yeast extract	5 g/L
SP supplements	1 g/L
KH_2PO_4	~56 mM (used to adjust pH to 5.7)
Agar (Serva)	2%
Glucose	3%

All ingredients, except glucose, are mixed and autoclaved at 121 °C for 20 minutes. Glucose is mixed in after autoclaving. 35 mL is used to fill a petri dish (Ø 8.5 cm) for standard usage or 50 mL is used to fill a plate used for genetic screens (Singer, PLU-003).

Table 9 YES 5-FOA plates

Compound	Final concentration
Yeast extract	5 g/L
SP supplements	1 g/L
KH_2PO_4	~56 mM (used to adjust pH to 5.7)
Agar (Serva)	2%
Glucose	3%
5-FOA	0.5 g/L

All ingredients except glucose and 5-FOA are mixed and autoclaved at 121 °C for 20 minutes. 5-FOA is dissolved in 1 L ddH_2O (60 °C). 5-FOA and glucose are added after autoclaving. 50 mL is used to prepare a plate used for genetic screens (Singer, PLU-003). Note that the concentration of 5-FOA is different from the 5-FOA concentration that is commonly used (1 g/L).

Table 10 YES antibiotic plates

Compound	Final concentration
Yeast extract	5 g/L
SP supplements	1 g/L
KH_2PO_4	~56 mM (used to adjust pH to 5.7)
Agar (Serva)	2%
Glucose	3%
5-FOA	0.5 g/L
Antibiotic	Depending on antibiotic, see below

All ingredients, except glucose and the antibiotic, are mixed and autoclaved at 121 °C for 20 minutes. Glucose and antibiotic are mixed in after autoclaving. Depending on the *S. pombe* strain that is to be grown, different antibiotics are used namely Geneticin (100 mg/L), Nourseothricin (100 mg/L) and hygromycin B (100 mg/L). For double or triple selection media multiple antibiotics are used. 35 mL is used to fill a petri dish (Ø 8.5 cm) or 50 mL is used to fill a plate used for genetic screens (Singer, PLU-003).

Table 11 SPAS plates

Compound	Final concentration
SP supplements	1 g/L
KH_2PO_4	~56 mM (used to adjust pH to 5.7)
Agar (Serva)	2%
Glucose	3%
1000x Vitamin mix*	1 mL/L

All ingredients, except glucose, are mixed and autoclaved at 121 °C for 20 minutes. Glucose is mixed in after autoclaving. 35 mL is used to fill a petri dish (Ø 8.5 cm) or 50 mL is used to fill a plate used for genetic screens. *1000x Vitamin mix contains pantothenic acid (4.2 mM), nicotinic acid (81.2 mM), inositol (55.5 mM) and biotin (40.8 mM).

2.2.2 *S. pombe* culturing

Unless otherwise mentioned *S. pombe* was grown in standard laboratory conditions i.e. in YES media at 30 °C (Hagan et al. 2016). For detailed media ingredients see Table 7 and Table 8. Generally, *S. pombe* was patched (~1 cm^2) from long-term storage on a YES

plate and left to grow at standard conditions for 2 days before a liquid culture was started. When *S. pombe* was grown in test tubes (2-10 mL), the liquid culture was rotated at medium speed. When *S. pombe* was grown in flasks (25-500 mL), the liquid culture was shaken a 150 RPM.

2.2.3 Long-term storage of *S. pombe* strains

For long-term storage *S. pombe* strains were plated on a YES plate and grown for 2 days at standard conditions. ~25 cm^2 of cells was collected and resuspended in freezing media (liquid YES media containing 30% glycerol). Glycerol was allowed to enter the cells for 10 minutes before long-term storage at -80 °C.

2.3 *S. pombe* mutagenesis

2.3.1 Standard *S. pombe* mutagenesis

2.3.1.1 Preparation of *S. pombe* genomic DNA

S. pombe was grown on plate at standard conditions. ~1 cm^2 of cells was collected in a 1.5 mL reaction tube. 200 mL of Smash and Grab buffer (2% (v/v) Triton X-100, 1% (v/v) sodium dodecyl sulfate, 100 mM NaCl, 10 mM Tris/HCl, pH 8.0, 1 mM EDTA, pH 8.0) was added along with 200 µL equilibrated phenol and 250 µl of Zirconia/Glass-Beads. Cells were lysed by vortexing for 3 minutes, quickly spun down before 200 µL ddH_2O was added. Samples were centrifuged at 16.000 g for 5 minutes. 350 µL of the upper phase was moved to a fresh reaction tube where 1 mL of ice-cold (-20 °C) 100% ethanol was added. After mixing, the sample was centrifuged at 16.000 g for 5 minutes. Supernatant was removed and 500 µL of ice-cold (-20 °C) 70% ethanol was added to the pellet. After mixing, the sample was centrifuged at 16.000 g for 5 minutes. Again the supernatant was removed, after which the pellet was dried for 5 minutes at room tempera-

ture (RT) in a vacuum concentrator. Finally, the pellet was dissolved in 500 µL of water. 1 µL of this was used in subsequent polymerase chain reactions (PCRs) reactions.

2.3.1.2 Preparation of PCR product for *S. pombe* transformation

To introduce mutations in the *S. pombe* genome, for example gene deletions, *S. pombe* was transformed with a DNA fragment displaying 80-500 bp homology up- and downstream of the region to be altered. This fragment was amplified by PCR using a DNA template that conveys resistance to geneticin, nourseothricin or hygromycin B, which was used to select mutants. If applicable, the PCR product also contains additional the DNA sequences that were introduced.

If the mutated allele already exists in a different strain of the *S. pombe* collection of the Braun Lab (including the haploid deletion mutant library (Bioneer, version 3.0)) a primer pair was designed to bind ~500 bp up- and downsteam of the mutation. When a full gene was deleted primers were designed using an inhouse perl script (Sigurd Braun, Ludwig-Maximilians-Universität München). If the mutated allele was not present in the *S. pombe* strain collection of the Braun Lab, primers are designed to bind a pFA6 plasmid with appropriate selection marker and have 80 bp of homology up- and downstream of the region that was mutated. The PCR reaction was performed using Robust (KAPA2G, KK5005) or Verify (PCRBIO, PB10.42-05) polymerase according to manufacturer's instruction.

2.3.1.3 *S. pombe* transformation

For transformation *S. pombe* was grown in 50 mL of liquid YES media under standard conditions, untill mid log phase (OD_{600} 0.4-0.8). Cells were pelleted at 700 g at RT and. washed once with 10 mL water and once with 5 mL LiOAc/TE solution (1 M LiOAc, 1

M Tris/HCl pH 8.0, 0.5 M EDTA pH 8.0). Cells were then resuspended in 1 mL Li-OAC/TE solution.

For transformation 100 µL of resuspended cells were used. 0.5-10 µg (generally 25 µL) of PCR product for *S. pombe* transformation (see section 2.3.1.2) was added along with 10 µL salmon sperm carrier DNA (10 mg/mL)(Sigma, D1626-1g). This was vortexed and incubated at RT for 15 minutes. 500 uL of PEG/LiOAc solution (50% PEG, 1 M LiOAc, 1 M Tris/HCl pH 8.0, 0.5 M EDTA) was added and sample was vortexed again and incubated at 30 °C for 50 minutes while nutated at 500 RPM. 50 µL DMSO was added to the sample after which the sample was vortexed. Cells were heat shocked at 42 °C for 10 minutes. The cells were centrifuged for 3 minutes at 2000 RPM, supernatant was removed and cells were resuspended in 100 µL YES and plated on YES plates, to recover. Cells were allowed to grow for 2 day at 30 °C after which cells are replica plated on selective media. Single colonies were picked and patched on selective media. Correct insertion of the mutant was confirmed by diagnostic colony PCR.

2.3.1.4 Diagnostic colony PCR in *S. pombe*

To verify if genomic integrations are correct, a pipette tip-sized patch of *S. pombe* was resuspended in zymo solution (2.5 mg/mL zymolyase (Seikagaki America, Inc., 120493-1), 0.1 M sodium phosphate buffer pH 7.5). Samples were heated to 37 ºC for 20 minutes followed by 95º C for 5 minutes. Sample was diluted by adding 150 µL of water, 1 µL of the resulting sample was used for diagnostic PCR.

Diagnostic PCR was performed using *Fast* polymerase (Nippon, KK5103) according to manufacturer's instructions. Generally, three primer pairs were used to verify the correct insertion or deletion: a primer pair examining the 5' integration (primer ~500 bp upstream of insertion and primer in the insert), a primer pair examining the 3' integration (primer in

the insert and primer ~500 bp downstream of insertion) and a primer pair testing for the deletion of the gene (two primers annealing within the ORF). In case a full gene was deleted primers were designed using an inhouse perl script (Sigurd Braun, Ludwig-Maximilians-Universität München).

2.3.2 Exceptions to the standard *S. pombe* mutagenesis

Several strains described in this work were created using methods different to the ones described above. These alternative strategies are described here.

2.3.2.1 Ectopic TAS strain generation

To generate a strain with ectopically inserted TAS at the *leu1*$^+$ locus 790 bp of the TAS1 region was amplified from genomic DNA obtained from strain PSB0017 as described in section 2.3.1.1, using oligonucleotides Sg3182 and Sg3183 (sequences: GGGCCCCCCCTCGAGGTCGACTATTTCTTTATTCAACTTACCGCACTTC and CGCTCTAGAACTAGTGGATCCGATGAATGGATTAAAAGGTGTTGG). The PCR product was purified and cloned into the pJK148 plasmid (ESB508). 10 µg of the pJK148 plasmid containing the TAS1 was linearized by NdeI (NEB, R0111S) digestion (2 hours, according to manufacturer's instruction) and transformed into *S. pombe* according to section 2.3.1.3. For transformation, the ST3479 strain (gift by J. Kanoh, Osaka University) was used, which has no endogenous TAS regions.

2.3.2.2 8xLTR reporter strain generation

To generate the 8xLTR-ura4 reporter strain, first using a CRISPR-Cas9 strategy (described in section 2.4) the endogenous *ura4*$^+$ gene was deleted from strain PSB0006, resulting in PSB2797. The hygromycin B resistance cassette was amplified from pFA6a-hygMX6 using primers Sg3860 and Sg3861 (sequences: GGTCTGAGTAGAAATTT-

GCTGCGGCATCCTTTCCTTTTATATTTGAAAAACCCCTCAAGT AAC-GAATGGTCATGTGACACGACATGGAGGCCCAGAATAC and TAGCG-GACTCGTTACTTACGTTGCGAGGTGGCTCAGGTGCAAGAAAGCCACCTCCCT GACCCACTCACGACGCTAAATCCCAGTATAGCGACCAGCATTCAC) and inserted ~4 kb from the *leu1*$^+$ gene in PSB2797. Next the 8xLTR reporter cassette was prepared on a pBluescript ii sk+ plasmid (ESB569). To this end LTR039 was amplified with unique primers Sg3800 and Sg3801 (sequence: TTTGCATGGGGATAGGTCGC and TCTGCCTTCAGCTACTTGGC) that bind up- and downsteam of the locus. The PCR product was purified using a PCR cleanup kit (Metabion, mi-PCR250). The cleaned up PCR product was amplified using primers Sg3788 and Sg3789 (sequences: TTT-GCATGGGGATAGGTCGC and AAAGTCGACTGTTTGCTACACAGTTTGG) that bind LTR039. This 2-step amplification ensured unique amplification of LTR039. The PCR product was cloned into pBluescript ii sk+ plasmid (ESB569) using KpnI and SalI digestion (NEB, R3142S and R3138S) and ligation (NEB, M0202S) according to manufacturer's instructions. This gave rise to plasmid ESB583. To prepare this plasmid for insertion into the *leu1*$^+$ locus 5' and 3' *leu1*$^+$ homology regions were sequentially cloned into the plasmid. The *leu1*$^+$ homology regions were amplified from the same genomic DNA as LTR039 using primers Sg3786 and Sg3787 (sequences: TTTCTCGAGTTT-GAAGAACAC and AAAGTCGACCAGTGCGCTCACCAAAGTAAC) and primers Sg3798 and Sg3799 (sequences: TTTCTCGAGAGGACAACGGATCGGGTTATGC and AAAGTCGACGAAGCAGATAAAATTGTACC). The PCR products were cloned into the ESB583 using XhoI and SalI digestion (NEB, R0146S and R3138S) and ligation (NEB, M0202S) according to manufacturer's instructions.

Next the LTR039 was "doubled" as described in Robinett et al. (1996)(Figure 2.1). Next the 2xLTR039 was doubled again using the same method. Then an *ura4*$^+$ reporter gene

was inserted on a pBluescript ii sk+ plasmid (ESB569) in a similar fashion as the LTR039 using primers Sg3796 and Sg3797 (sequences: TTTCTCGAG-TATCATCATAGCTTAGC and AAAGTCGACCTACATACATCTTTCATTGG). The *ura4*$^{+}$ reporter gene was placed downstream of the 4xLTR039 and downstream of this an additional 4xLTR039 cassette was inserted, both times using the same "doubling" strategy as used before. The resulting plasmid was digested with KpnI and SacII (NEB, R3142S and R3156S) according to manufacturer's instructions and the digestion product was transformed into PSB2814.

2.4 Mutant generation with CRISPR-Cas9

Genomic mutations using the CRISPR-Cas9 system were made as previously described (Torres-Garcia et al. 2020). In short, a guide RNA sequence was designed using the CRISPR4P tool (http://bahlerweb.cs.ucl.ac.uk/cgi-bin/crispr4p/webapp.py, (Rodríguez-López et al. 2016)). The guide RNA was complemented by flanking it with CTA-GAGGTCTCGGACT and GTTTCGAGACCCTTCC nucleotides at its 5' and 3' respectively. The resulting sequence and its reverse complement were ordered as oligonucleotides. To produce sgRNA fragment 5 µL of each oligonucleotide (100 µM) is mixed, heated to 95 °C for 3 minutes and allowed to cool at a rate of -2 °C/minute. The resulting sgRNA fragment was cleaned up (Monarch DNA Cleanup Kit (NEB, T1030S)). The diluted DNA was cloned into pLSB (gift form R. Allshire, Edinburgh University) by Golden Gate cloning (0.5 µL pLSB plasmid (75 ng/µL), 0.5 µL sgRNA fragment (1 ng/µL), 1 µL T4 DNA ligase Buffer (10x), 0.5 µL NEB Golden Gate Assembly Mix (NEB, E1601S) and 7.5 µL water, incubated at 37 °C for 1 hour followed by 5 minutes at 60 °C). Resulting plasmid was checked by an analytical digest with NcoI (NEB, R3193S, 1 hour according to manufacturer's instruction) and sequenced to check for correct inser-

tion of the guide RNA. 200 ng of the resulting plasmid is transformed together with 1000 ng of HR template (PCR product of 2 primers designed using the CRISPR4P tool) as described in section 2.3.1.3 with the exception that the cells are plated on YES plates and replica plated after 16 hours onto selective media, on which they are allowed to grow for 2 days. Colonies are streaked on YES plates to lose the pLSB plasmid. Colonies can be checked as described in section 2.3.1.4, using primers designed using the CRISPR4P tool.

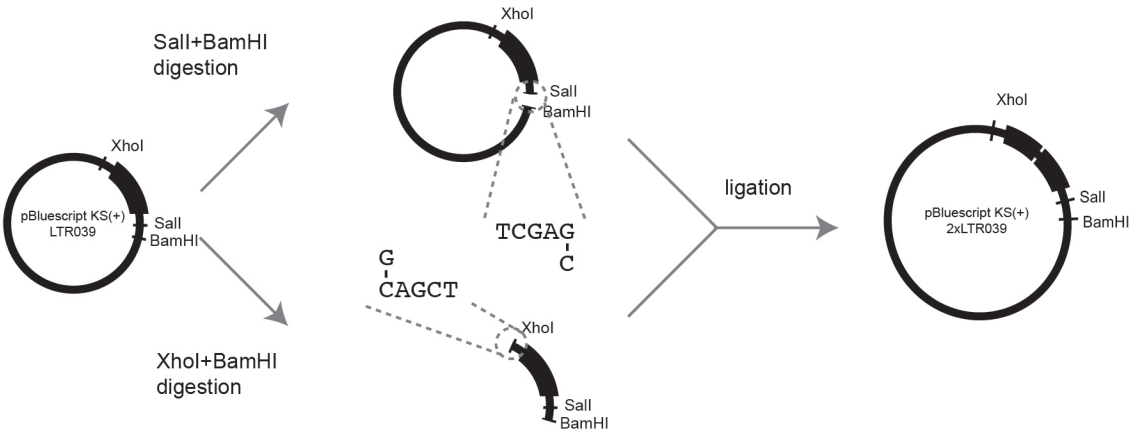


Figure 2.1 Strategy to clone repetitive LTR arrays

To double a LTR without the use of PCR the LTR is cloned into the multiple cloning site of a pBluescript ii sk+ plasmid. The resulting plasmids is digested with two combinations of enzymes SalI + BamHI and XhoI + BamHI. This generates fragments with compatible overhangs that can be ligated to form a pBluescript ii sk+ plasmid with a duplication of the original LTR. This process can be repeated to further expand the copy number of the LTR as SalI and XhoI digestions produce compatible overhangs but upon ligation both recognition sites are destroyed (Robinett et al. 1996).

2.5 *S. pombe* strains used in this dissertation

Table 12 *S. pombe* strains used in this dissertation

Strain number*	Genotype	Published/ created by	Figure
Bioneer library	*P (h+) leu1-32 ade6-M216 ura4-D18 geneX::kanMX*	(Kim et al. 2010)	Figure 3.18
PSB0006	*P (h-)*	(Bühler et al. 2006)	Figure 3.18
PSB0017 (FY1193)	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+*	(Ekwall et al. 1999)	Figure 3.2 Figure 3.3 Figure 3.6 Figure 3.8 Figure 3.9

PSB0023 (SBP005)	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ clr1::kanMX*	(Braun et al. 2011)	Figure 3.9
PSB0025 (SBP007)	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ mit1::kanMX*	(Braun et al. 2011)	Figure 3.8 Figure 3.9
PSB0065	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+*	(Barrales et al. 2016)	Figure 3.12 Figure 3.13 Figure 3.15 Figure 3.16 Figure 3.17
PSB0068 (SPT981)	*P (h+) leu1-32 ade6-210 otr1::ura4+ mit1-K587A*	(Sugiyama et al. 2007)	Figure 3.8
PSB0069 (SPT981)	*P (h+) leu1-32 ade6-210 otr1::ura4+*	(Sugiyama et al. 2007)	Figure 3.8
PSB0072 (FY1862)	*P (h90) leu1-32 his3D1 ade6-210 ura4-D18 otr1R(Sph1)::ade6+ tel(1L)::his3+ tel(2L)::ura4+*	(Nimmo et al. 1998)	Figure 3.3 Figure 3.11
PSB0072 (FY1862)	*P (h90) leu1-32 his3D1 ade6-210 ura4-D18 otr1R(Sph1)::ade6+ tel(1L)::his3+ tel(2L)::ura4+*	(Nimmo et al. 1998)	Figure 3.10
PSB0074 (PM0304)	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ rik1::natMX*	(Braun et al. 2011)	Figure 3.8 Figure 3.9
PSB0078	*P (h90) leu1-32 his3D1 ade6-210 ura4-D18 otr1R(Sph1)::ade6+ tel(1L)::his3+ tel(2L)::ura4+ rik1::natMX*	(van Emden et al. 2019)	Figure 3.11
PSB0090	*P (h+) leu1-32 ade6-210 ura4-DS/E(ura4-D18?) imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ clr4::kanMX*	(Braun et al. 2011)	Figure 3.16
PSB0374	*P (h+) leu1-32 ade6-M216 ura4-D18*	(Kim et al. 2010)	Figure 3.13 Figure 3.15
PSB0392	*P (h+) leu1-32 ade6-M216 ura4-D18 cbp1::kanMX*	(Kim et al. 2010)	Figure 3.13
PSB0565	*P (h-) leu1-32 ade6-M210 ura4-D18 smt0 locus2::mCherry::HygR mat3M::ura4+*	Ramon Barrales	Figure 3.18
PSB0580	*P (h-) leu1-32 ade6-210 ura4 imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ hygR::cen1*	Ramon Barrales	Figure 3.18
PSB0906	*P (h+) leu1-32 ade6-210 ura4 imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ hygR::cen1 lem2::natMX*	Ramon Barrales	Figure 3.13 Figure 3.15 Figure 3.16 Figure 3.17
PSB1022	*P (h-) leu1-32 ade6-210 ura4 imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ hygR::cen1 lem2::natMX h-SPSQ(cyhR) SPL42(cyhS) pREP81x_Lem2_GFP*	Ramon Barrales	Figure 3.16
PSB1480	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ clr3::kanMX*	Zsuzsa Sarkadi	Figure 3.13
PSB1487	*P (h+) leu1-32 ade6-M216 ura4-D18 clr4::kanMX*	(Kim et al. 2010)	Figure 3.13
PSB1488	*P (h+) leu1-32 ade6-M216 ura4-D18 clr3::kanMX*	(Kim et al. 2010)	Figure 3.13
PSB1525	*P (h+) leu1-32 ade6-210 ura4-DS/E(ura4-D18?) imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ clr4::kanMX lem2::natMX*	Sabine Stöcker	Figure 3.16
PSB1613	*P (h90) leu1-32 ade6-210 ura4-D18 his3D1 otr1R(Sph1)::ade6+ tel(1L)::his3+ tel(2L)::ura4+::hphMX*	Ramon Barrales	Figure 3.18

PSB1728	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ ccq1::natMX*	(van Emden et al. 2019)	Figure 3.3 Figure 3.6 Figure 3.8 Figure 3.9
PSB1730	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ clr3::kanMX*	(van Emden et al. 2019)	Figure 3.8 Figure 3.9
PSB1741	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ rik1::natMX ccq1::kanMX*	(van Emden et al. 2019)	Figure 3.8 Figure 3.9
PSB1780	*P (h+) leu1-32 ade6-210 ura4 imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ hygR::cen1 lem2::natMX red1Δ::kanMX*	Ramon Barrales	Figure 3.13
PSB1784	*P (h+) leu1-32 ade6-210 ura4 imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ hygR::cen1 red1Δ::kanMX*	Sabine Stöcker	Figure 3.13
PSB1786	*P (h+) leu1-32 ade6-210 ura4 imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ hygR::cen1 pab2Δ::kanMX*	Sabine Stöcker	Figure 3.13
PSB1788	*leu1-32 ade6-210 ura4-D18 IR-L(HpaI)::ade6+ mat1-M-smt0 his2 SPBC18E5.09c-LTR17::ura4+*	Sabine Stöcker	Figure 3.18
PSB2028	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ mit1::kanMX ccq1::natMX*	(van Emden et al. 2019)	Figure 3.8 Figure 3.9
PSB2029	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ clr3::kanMX ccq1::natMX*	(van Emden et al. 2019)	Figure 3.8 Figure 3.9
PSB2082	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ mit1::kanMX taz1::hygMX*	(van Emden et al. 2019)	Figure 3.6
PSB2083	*P (h+) leu1-32 ade6-210 Ura4-DS/E imr1L(NcoI)::ura4+otr1R(SphI)::ade6+ taz1::hygMX ccq1::kanMX*	(van Emden et al. 2019)	Figure 3.6
PSB2102 (K240)	*M (h-) leu1-32*	(Yamada et al. 2008)	Figure 3.5
PSB2103 (KYP176)	*M (h-) leu1-32 snf21-36(ts)*	(Yamada et al. 2008)	Figure 3.5
PSB2127 (FY520)	*Mat1Msmt0 leu1-32 his2- ura4 DS/E ade6-210 m23::ura4-Tel72*	(Nimmo et al. 1994)	Figure 3.3
PSB2127 (FY520)	*Mat1Msmt0 leu1-32 his2- ura4 DS/E ade6-210 m23::ura4-Tel72*	(Nimmo et al. 1994)	Figure 3.10
PSB2314	*P (h+) leu1-32 ade6-M216 ura4-D18 lem2::kanMX*	This study	Figure 3.13 Figure 3.15
PSB2336	*P (h90) leu1-32 his3D1 ade6-210 ura4-D18 otr1R(Sph1)::ade6+ tel(1L)::his3+ tel(2L)::ura4+ clr3::natMX*	(van Emden et al. 2019)	Figure 3.11
PSB2338	*M (h-) ade6-M216 his7-366 leu1-32 ura4-D18 SH1L::ura4+ SH1R::his7+ SH2L::his7+ SH2R::his7+ SH3L::ura4+ Leu1::Leu1-TAS1(800bp)*	(van Emden et al. 2019)	Figure 3.3
PSB2369	*M (h-) ade6-M216 his7-366 leu1-32 ura4-D18 SH1L::ura4+ SH1R::his7+ SH2L::his7+ SH2R::his7+ SH3L::ura4+ Leu1::Leu1-TAS1(800bp) ccq1::natMX*	(van Emden et al. 2019)	Figure 3.3
PSB2371	*P (h90) leu1-32 his3D1 ade6-210 ura4-D18 otr1R(Sph1)::ade6+ tel(1L)::his3+ tel(2L)::ura4+ clr3::natMX*	(van Emden et al. 2019)	Figure 3.11
PSB2684	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ set1::kanMX*	This study	Figure 3.13

PSB2686	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ ago1::KanMX*	This study	Figure 3.13
PSB2701	*P (h+) leu1-32 ade6-210 ura4-DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ hip1::kanMX*	This study	Figure 3.13
PSB2797	*P (h-) ura4Δ*	This study	Figure 3.18
PSB2814	*P (h-) ura4Δ SPBC1E8.05:: SPBC1E8.05-hphMX*	This study	Figure 3.18
PSB2860	*P (h-) ura4Δ SPBC1E8.05:: SPBC1E8.05-hphMX leu1::4xLTR-ura4-4xLTR*	This dissertation	Figure 3.18
PSB2861	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ wdr7::kanMX*	This study	Figure 3.20
PSB2862	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ wdr7::kanMX lem2::natMX*	This study	Figure 3.20
PSB2863	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ sft2::kanMX*	This study	Figure 3.20
PSB2864	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ sft2::kanMX lem2::natMX*	This study	Figure 3.20
PSB2865	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ cid14::kanMX*	This study	Figure 3.20
PSB2866	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ cid14::kanMX lem2::natMX*	This study	Figure 3.20
PSB2867	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ vps26::kanMX*	This study	Figure 3.20
PSB2868	*P (h+) leu1-32 ade6-210 ura4DS/E imr1L(NcoI)::ura4+ otr1R(SphI)::ade6+ vps26::kanMX lem2::natMX*	This study	Figure 3.20

* Strain number in brackets denotes strain name in original laboratory.

2.6 Serial dilution growth assay (silencing assay)

To perform a serial dilution assay 5 mL YES was inoculated with a *S. pombe* culture, which was grown O/N in standard conditions. The O/N culture was diluted 1:10 into 1.5 mL of YES; 1 mL of diluted culture was used to measure the OD_{600}. YES was used to dilute the remaining culture to an OD_{600} of 0.3. This culture was serially diluted in 1:3 steps in a 96 well plate. Using a sterilized pin frogger (pin diameter 0.3 cm), the cultures were plated on selective and non-selective media plates. These were grown for 2-3 days at 30 °C before pictures were taken.

2.7 Quantitative gene expression analysis

2.7.1 Reverse transcription

For RT-qPCR *S. pombe* was grown in 50 mL of liquid YES media as described above, till mid log phase (OD_{600} 0.4-0.8). Cells were pelleted at 700 g for 5 minutes and washed with 50 mL of cold (4 °C) water. Cells were pelleted once again and flash-frozen. To extract RNA the pellet was thawed, on ice, in 1 mL of TRIzol (Life Technologies, 15596018). 250 µl of Zirconia/Glass-Beads (Roth/Stricker, N034.1) was added and the cells were broken up by 3 rounds of 30 seconds beat beating (Peqlab, Precellys 24). Samples were centrifuged at 13500 RPM at 4 ºC for 10 minutes. Cleared lysated was added to 200 µL of chloroform (Sigma Aldrich, C2432-1L) and mix thoroughly by vortexing for 15 seconds, immediately after addition. RNA was extracted by centrifuging at 12,000 g at 4 ºC for 10 min. The aqueous phase was added to 500 µL of chloroform, mixed and again centrifuged at 13500 RPM at 4 ºC for 10 min. The aqueous phase was added to 500 µL isopropanol briefly vortexed and incubated at 4 ºC for 15 minutes after which the sample was centrifuged at 12,000 g x at 4 ºC for 10 minutes. Supernatant was removed and pellet was washed twice with ice cold 75% ethanol and centrifuged at 9600 RPM for 5 min at RT. Supernatant was removed and pellets were dried in a vacuum concentrator for 5 minutes. RNA was resuspended in 100 µL RNase-free water (Thermo Fisher, 10977035). The RNA was DNase treated by diluting 20 ng of RNA in 36 µL RNase-free water. 4 µL of TURBO DNA-free 10x buffer and 0.5 µL TURBO DNA-free DNase I (Ambion, M1907) was added to the sample which is then incubated at 37 ºC for 30 minutes. Another 0.5 µL TURBO DNA-free DNase I was added for another 30 minutes. To inactive the DNase I, TURBO DNase inactivation reagent was added and sample was mixed thoroughly and incubated at RT for 2 minutes. During the RT incubation the sample is mixed occasionally. Sample was centrifuged 10,000 g for 1.5 minutes, 20-30 µL of the superna-

tant were transferred to a fresh tube. Next 11 µL of DNase treated RNA was mixed with 1 µL of 50 µM oligo-$(dT)_{17}$ primer and 1 µL of 10 mM dNTP mix (Roche, KK1007). For denaturing (i.e. removal of secondary structures), the sample was incubated at 70 ºC for 10 minutes after which the sample was incubated on ice for 10 minutes. For cDNA synthesis denatured DNase treated RNA was incubated with 4 µL 5x First-Strand Buffer (Thermo Fisher, 18080085), 1 µL 0.1 M DTT (Sigma, 43819-25G), 1 µL RNasin (homemade), 0.25 µL SuperScript III (Thermo Fisher, 18080085) and 0.75 µL water at 50 ºC for 1 hour and heat-inactivated at 70 ºC for 15 minutes. To prepare samples for qPCR the cDNA was diluted to 1:25 for heterochromatin targets and 1:1000 for euchromatin targets. cDNA quantification was done as described in section 2.9.

RT-qPCR experiments related to the TAS regions are pool-normalized i.e. independent experiments are standardized to the actin normalized mean of all samples from each experiment. Data is shown relative to the pool-normalized wild type. RT-qPCR experiments related to LTRs are not pool-normalized, instead individual experiments are normalized to their corresponding actin signal and shown relative to wild-type.

2.7.2 RNA-seq

cDNA was made as described in section 2.7.1. Libraries were made using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7760L) according to manufacturers instructions. 50 bp, single-end reads were sequenced at the Illumina HiSeq 1500 sequencing platform. For detailed processing of the reads see section 2.13.5.

2.8 Chromatin immunoprecipitation methods

2.8.1 ChIP-qPCR

For ChIP-qPCR *S. pombe* was grown in 100 mL of liquid YES media at standard conditions, till mid log phase (OD_{600} 0.4-0.8), except in case of the Lem2-GFP ChIP-qPCR experiment where the cells were grown in EMM –Leu media to maintain the pREP81x_Lem2_GFP plasmid. Cells were crosslinked in the presence of a final concentration of 1% formaldehyde (Roth, 4979.1) for 10 minutes at RT. Cells were pelleted at 700 g for 5 minutes and washed twice with 50 mL of PBS (10 mM Na-phosphate, 137 mM NaCl, 2.7 mM KCl, pH 7.4 adjusted with HCl) buffer. Cells were pelleted once again and flash-frozen. For performing the immunoprecipitation, the frozen pellet was first thawed on ice in 500 mL of lysis buffer (50mM HEPES/KOH pH7.5, 140 mM NaCl, 1 mM EDTA, 1% TritonX-100, 0.1% NA-DeoXycholate). Then 250 µL of Zirconia/Glass-Beads (Roth/Stricker, N034.1) were added and the cells were broken up by 4 rounds of 30 seconds beat beating (Peqlab, Precellys 24). Chromatin was separated from beads and cell debris by a short (5 second, 5000 RPM) centrifugation step. Chromatin was sheared by sonication for a total of 30 minutes using 30 second on/off cycles (Qsonica, Q800R). For immunoprecipitation chromatin was incubated O/N with antibody at 4 °C with nutation; for antibody amounts and manufacturer see Table 13. Hybrids were captured by 4 hour incubation with ProtG-Dynabeads (Invitrogen, 10004D). Beads were washed twice with Lysis buffer, twice with high salt buffer (50 mM HEPES/KOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% TritonX-100, 0.1% NA-DeoXycholate) and twice with wash buffer (10 mM Tris/HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% TritonX-100, 0.5% Na-DeoXycholate). Each wash step is 1 minute while the samples are nutated, samples a separated from the wash buffers using a magnetic rack. To reverse the crosslinks, samples were incubated for 4 hours at 65 °C and afterwards treated for 1.5 hours with 10

µL ProtK (4 mg/mL) at 55 °C. Samples were purified using a kit (Zymo Research, D5201).

To prepare samples for qPCR they were diluted 1:25 and analyzed along-side the input (1:100 diluted) and, in certain cases, also a no-antibody control (diluted 1:25). Samples were quantified as described in section 2.9.

For the data analysis mean and Standard Error of the Mean values were calculated from three to four independent experiments unless otherwise noted. To avoid bias against under- or overrepresented sequences samples were normalized against input. Unless otherwise noted input-normalized qPCR signals were furthermore normalized to an internal control using the mean of three euchromatic loci (*act1*$^+$, *ade2*$^+$, *tef3*$^+$) to generate "EC normalized" values.

2.8.2 Antibodies used for ChIP

Table 13 Antibodies used in ChIP experiments

Antigen	Manufacturer	Amount used in ChIP
H3	Active Motif (61475)	2 µL
H3K9Me2	Abcam, (ab1220)	2 µL
GFP	In house serum	2 µL
RNA PolII-Ser5Phos	In house serum	25 µL

2.8.3 ChIP-seq

For ChIP-seq immunoprecipitated DNA was prepared as for ChIP-qPCR. Libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumnina (NEB, E7645S) according to manufacturer's instruction with pooled samples of three independent experiments with a total amount of cells corresponding to 120 OD_{600}. 50 bp single-end reads were sequenced at the Illumina HiSeq 1500 sequencing platform. Mapping of Illumina reads and normalization were performed as previously described (mapping and normalization performed by Cornelia Brönner)(Brönner et al. 2017). Reads were mapped

to the *S. pom*be genome (ASM294v2.27) supplemented with the TAS sequence from the pNSU70 plasmid (Sugawara 1988).

2.9 Quantitative PCR

For quantitative PCR (qPCR) a SYBR Green Master mix (Life Technologies, A25778, Steinbrenner-laborsysteme, SL-9904-50mL, Thermo Fisher Scientific, A46109) was used according to manufacturer's instructions. Reaction volume (3 parts SYBR green Master mix, 2 parts cDNA or ChIP sample and 1 part primer pair (0.25 µM per primer)) for 96 well plate was 15 µL and reaction volume for 384 well plate was 10 µL. Samples were quantified using the 7500 Fast real-time PCR system, the QuantStudio3 or QuantStudio5 (Applied Biosystems). Primers used are listed in Table 14.

2.9.1 Primers used for qPCR

Table 14 Primers used for qPCR

oligo number	FOR oligo	REV oligo	locus	reference
Sg0243/0244 (P059/060)	TGCTCTGACTTGGCTTGTCTT	CCCTAACTTGGAAAGGCACA	*cen-dg*	(Braun et al. 2011)
Sg0272/0273 (mb274/276)	ATGGTCGTCGCTTCAGAAATTGC	CTCCTTGGAAGAATTGCAAGCCTC	*tlh1*$^+$	(Bühler et al. 2007)
Sg0423/0424 (P638/639)	AACCCTCAGCTTTGGGTCTT	TTTGCATACGATCGGCAATA	*act1*$^+$	(Braun et al. 2011)
Sg1026/1027	CAGCAATATCGTACTCCTGAA	ATGCTGAGAAAGTCTTTGCTG	*ura4*$^+$	(van Emden et al. 2019)
Sg1038/1039	GAAGTTCACTCAGTCATAATTAATTGGGTAAC	GGGCCCAATAGTGGGGGCATTGTATTTGTG	TERRA	(Bah et al. 2012)
Sg1804/1806	CCGTCTATATACATTGATATCCGTTG	CTGCGGTGAGTTTTACTTGC	LTR098	This study
Sg1906/1097 (Telomeric	CGGCTGACGGGTGGGGCCCAATA	GTGTGGAATTGAGTATGGTGAA	20 bp from tel repeats	(Tomita & Cooper 2008)

STE1)				
Sg1908/ 1909 (JK380 /381)	TATTTCTTTATTCAACTTA CCGCACTTC	CAGTAGTGCAGTG-TATTATGA-TAATTAAAATGG	TAS1/ 100 bp from tel. repeats	(Harland et al. 2014)
SG1953 /1954	TCGCCGG-TAACAAAAGGATCA	GCATTAGA-CAACTCGTTCGATC	cnt1	(Barrales et al. 2016)
Sg2038/ 2039	TTATTCACCCATACACAC-TACACC	GATGAATGGAT-TAAAAGGTGTTGG	744 bp from tel repeats	(van Emden et al. 2019)
Sg2102/ 2103	ATCTACTCCAA-TATAGTCCTCTGC	GATAATGGATGGAGGTAA TAATGGATGGAGGTAA-GAGAGG	1297 bp from tel repeats	(van Emden et al. 2019)
Sg2106/ 2107	TTATATTCCTG-CATCCCAACACAT	AAAGAAGATAAAA-GCAGGGGACTA	2275 bp from tel repeats	(van Emden et al. 2019)
Sg2139/ 2149	TCGTTAACAACATTTAAC GATTACTCG	ACGTTTGTTGAGTGA-TATGTCGTCG	2758 bp from tel repeats	(van Emden et al. 2019)
Sg2141/ 2142	TACTCCAACACAC-TCAATACATACC	AAGTAGGAGAATGAA-GAAGTAATCAAAG	2012 bp from tel repeats	(van Emden et al. 2019)
Sg2176/ 2177	TCATTCATTCATTCAATTT GCAC	CTGTTGTAATTT-GTCGCTGAGA	LTR026	This study
Sg2671/ 2672	AG-GCATCTGATCCCAATGAG	ATTTTGGATGCCTT-GGATGA	*ade2*$^{+}$	(van Emden et al. 2019)
Sg2736/ 2737	TGGCCTTCTTAGCCTTTTC A	CTGAGGAAGTTT-GGGCTGTC	*tef3*$^{+}$	(van Emden et al. 2019)
Sg2983/ 2984	GGTAAGCCTAGTAAC-GATGCC	GTGCCAACAGTGATAC-GCAA	*his3*$^{+}$	(van Emden et al. 2019)
Sg2985/ 2986	CCCCGACGACATATCAC-TCA	ATGAACGACAAACAG-CAGGC	TAS2	(van Emden et al. 2019)
Sg2991/ 2992	CCTCTGACAGATGCTCAA ACC	TGGTTAC-GGTTATTAGGTGATGT	TAS3	(van Emden et al. 2019)
Sg3119/ 3120	CAATT-GGGCCGAATGATGGT	TGCTCAC-GTCCTCCATCAAT	*ade6*$^{+}$	(van Emden et al. 2019)
Sg3121/ 3122	TGACCCCGATGCAATT-GTTG	AGAGTTGCAGGA-GAGGGTTC	*ade6*$^{+}$	(van Emden et al. 2019)
Sg3123/ 3124	TTCCAG-TAATCGGCGTTCCT	CGACAGGCTAAAATAC-CGGC	*ade6*$^{+}$	(van Emden et al. 2019)
Sg3477/ 3478	ACGATGCATTGCGAT-TTTTGT	GAGGAATGAGGTTCAG-TTGTAGC	LTR061	This study
Sg3479/ 3480	GTAAGTGCAG-TCATTTATACACCTT	GGAACGAGGCTCGGTT-GTAT	LTR222	This study
Sg3481/ 3482	TGTCAAGTGCACTGTTTT-GATT	AAACCGAT-TCCCGTCCTCAC	LTR050	This study
Sg3496/ 3497	ACTGAACTGAGGAAC-GAGGT	TGTGGAATACAAGTTAAA-GAATACAAGTTAAA-GATCCG	LTR095	This study
Sg3511/	GCTACAACTGAATTTCGTT	ACAATGTAGCGG-	LTR123	This

3512	CCTCA	TATGCGGT		study
Sg3523/ 3524	AATCACGGGTAAC-TATTCGTCT	ACCTCGTTCCTCAG-TTCAGT	LTR204	This study
Sg3562/ 3564	GGAATACTAGCTAA-GATCCG	TCTAAATAGTT-GTTATCAACG	Tf1	This study
Sg3567/ 3568	TGTCAGCAATACTACAC-TACG	AACGAGGTTCAGCAG-TAGC	Tf2	This study
Sg3734/ 3735	ACTGCACCACACATTAG-TGA	ACTGAAATGAGAAAC-GAGGGTCA	LTR164	This study
Sg3736/ 3737	CCTCGTAGATGATAGA-TATTGGGGC	CTTCTGACTAAAGTGAG-GAACGA	LTR146	This study
Sg4322/ 4323	GCTACTTATACAAC-CTGTGTGTTGC	ACGAAATGTAAAATT-GCGGTGAGT	LTR179	This study
Sg4324/ 4325	ACTGCTAAAC-CTCATTCCTTACACA	TGCGGTAATTTTTCCTT-GTTAACTT	LTR174	This study
Sg4326/ 4327	TCCTTC-TACATGCATGGATTCCC	CTCAGTTCAGCTAAA-GCTCTATTTT	LTR235	This study
Sg4328/ 4329	TCCATGTTCCATTTACACA TCGT	AGTGAGGAGCACAG-TTCAGC	LTR096	This study
Sg4330/ 4331	TCACGGGTAAC-TATCCGTCTAT	ACTGAG-GAACAAGGTTCAGTTGT	LTR077	This study
Sg4332/ 4333	GCATACTTCCGGA-TAACTGCG	ACGAGATTCAAATT-GCGTTCTTTCT	LTR009	This study
Sg4334/ 4335	ACATGATAAGTCATCAA-GATGGAGA	ACTGAACTGAGGGAC-GAGGT	LTR118	This study
Sg4336/ 4337	ACAGGTTGTATAAGTAG-CAACTGGA	TGCTGAAC-CTCGTTCCTCAG	LTR126	This study
Sg4338/ 4339	TCACATGTTCTGTTATCTA TCCGCT	AC-GAATTCCAGCTAGGCTCT	LTR203	This study
Sg4340/ 4341	CTGTCTATACACAT-ACTCGCTGA	AGGAACGAGGTTCAGTT-GGAA	LTR125	This study
Sg4342/ 4343	TCGCTCCTCAGTTCAC-TTGT	GGCGACTTTGCTCATGG-TAAG	LTR114	This study
Sg4346/ 4347	TGGGAAGAACCAAAC-CTCGT	CACAGTGCCTTAATGTTT-GAGAA	LTR089	This study
Sg4348/ 4349	TTGCGGGTCACTAG-TTTCGT	GAGGAATGAGAC-GCAGCTGTA	LTR177	This study
Sg4350/ 4351	TTTGCCATGTAGCGG-TAGGT	ACAACTGAAC-CTCGCTCCTTT	LTR056	This study
Sg4354/ 4355	GCTACAACTAAAC-CTCGTCCCTA	TCTTGCGAAC-TACAATATATGCG	LTR157	This study
Sg4356/ 4357	TACGACTTACAA-TATAGCGGTATGC	GAGGAAC-GAGGTTCGGCTG	LTR121	This study
Sg4358/ 4359	TGCGATAGTAAC-GAACAACGA	TGAGGAATGAGGTTCAG-TGTGA	LTR216	This study
Sg4362/ 4363	ATCCCGC-TACAAAGTATACGAA	TGGTATTGC-TACAATCTGTCGT	LTR013	This study

2.10 Subtelomere stability assay

Telomere stability assays were performed with freshly prepared *ccq1*Δ cells, which were generated by homologous recombination as described in section 2.2. For other mutants, individual colonies were selected from freshly grown colonies from glycerol stocks (stored at -80ºC) as described in section 2.2.2. Individual mutants were used to inoculate

5 mL of YES media to OD_{600} ~ 0.05. These cultures were grown O/N to an OD_{600} of 3-5. 5 OD_{600} of cells were harvested by centrifugation and flash-frozen. The remaining culture was used to again set up 5 mL of YES media at OD_{600} ~ 0.05, which again were grown O/N and harvested. In total samples were collected over 6 days. A yeast DNA extraction kit (Thermo Fisher Scientific, 78870) was used to extract genomic DNA from frozen cell pellets. To prepare samples for qPCR they were diluted 1:100. Relative copy numbers of individual reporter genes or subtelomeric loci were measured by qPCR as described in section 2.9 and normalized to internal reference genes (*act1*$^+$, *ade2*$^+$and *tef3*$^+$).

In order to quantify the loss of the *ura4*$^+$ reporter from minichromosome Ch16, the propagation and stability of the Ch16 minichromosome itself has to be taken into account as well. This was done by quantifying the qPCR signal of *ade6*$^+$, this gene is present on chromosome 3 and on the Ch16 minichromosome. Since the Ch16 minichromosome and *ura4*$^+$ have equal levels at the start of the experiment, the contribution of genomic (Chr III) *ade6*$^+$ to the total *ade6*$^+$ signal can be determined by subtracting the qPCR signal of *ura4*$^+$ from qPCR signal of *ade6*$^+$. The calculated signal for the genomic (ChrIII) *ade6*$^+$ remains constant throughout the experiment whereas the qPCR signal for minichromosomal *ade6*$^+$ decreases with loss of the Ch16 minichromosome. This loss can be quantified by subtracting the constant genomic (ChrIII) *ade6*$^+$ from the total *ade6*$^+$ signal for each experimental day. The calculated minichromosomal loss was then used to normalize the *ura4*$^+$ reporter copy number.

2.11 Telomere-PCR

Telomere-PCR was performed essentially as described in (Moravec et al. 2016). Genomic DNA was denatured and poly(C)-tailed. Telomeres were amplified in a PCR reaction with primers that recognize the poly(C) tail and a downstream region. To analyze the telomere

length the PCR products were analyzed by running them on a 0.8% agarose gel. Primers used for Telomere-PCR are Sg2708 and Sg2709 (sequences: TGAGTGTGCTGGAG-TACGTT and CGGGATCCGGGGGGGGGGGGGGGG).

2.12 Genome-wide genetic screen for LTR regulators

To perform a genetic screen a haploid deletion mutant library (Bioneer, version 3.0) was crossed with a reporter strain (for strain construction see section 2.3) using a high-density array pinning robot (Singer, ROTOR HDA). For the first step the library and the reporter strains were seeded in 384 format on YES media plates with the appropriate selection markers (geneticin for the haploid deletion mutant library, hygromycin B for the reporter strain). After two days of growth at 32 °C the library and reporter strain were crossed by plating them on plates containing SPAS media. For more efficient mating a drop of water was added. Library and reporter strain were allowed to mate by incubating them for 3 days at; the plates were then moved to 42 °C for 4 days to kill all parental cells that have not mated. Spores were plated on selective YES media and allowed to germinate at 32 °C for 2 days. At this step, from each cross, the 2 copies were produced (each patch was replica plated onto selective media twice). After three days the selection was repeated. Two days later the silencing assay was started by plating the 2 copies on YES and 0.5 g/L 5-FOA YES media plate. Cells were grown at 32 °C and pictures were taken every 24 hours.

2.13 Computational biology methods

The custom genome sequence and annotation file, along with all code to perform the computational methods described in this section can be found at https://github.com/tsvanemden/Thesis.

2.13.1 Genomic sequences

The *S. pombe* genome sequence (ASM294v2.27) and its annotation file were downloaded from https://www.pombase.org/datasets. The sequence of the pNSU70 plasmid was downloaded from https://www.pombase.org/status/telomeres.

To study the TAS region and neighboring subtelomeric sequences the *S. pombe* genome (ASM294v2.27) was complemented with pNSU70 based on homology (work performed by Cornelia Brönner).

To study LTRs the *S. pombe* genome sequence (ASM294v2.27) was used, the corresponding annotation file was complemented with 239 LTRs found in the EMBL format genome sequence and features (https://www.pombase.org/downloads/genome-datasets) visualized with Artemis (version 16)(http://www.sanger.ac.uk/science/tools/artemis).

2.13.2 Dot matrix analysis

The dotmatrix analysis was performed in R using the dotplot function of the seqinr package (https://github.com/evolvedmicrobe/dotplot).

2.13.3 *In silico* nucleosome prediction

The *in silico* nucleosome occupancy prediction was performed in R using a the NuPoP package (Xi et al. 2010). The genomic sequence used in this analysis is the *S. pombe* genomic sequence complemented with the pNSU70 as described above.

2.13.4 *In silico* DNA sequence analysis

The *in silico* DNA sequence analysis was performed using Python. The genomic sequence used in this analysis is the *S. pombe* genomic sequence complemented with the pNSU70 as described above.

2.13.5 RNA-seq analysis

NGS reads were mapped to the *S. pom*be genome (ASM294v2.27). The annotation file supplemented with the coordinates of the LTRs was used to quantify reads at LTRs. Because LTRs are repetitive assigning reads to individual LTRs is difficult. Reads were processed in 3 different ways to test behavior of the algorithms used: (1) Either only uniquely mapping reads were taken into account, (2) randomly selected alignments of multi-mappers were taken into account, or (3) expectation maximization was used to quanify transcripts (Figure 2.2). Reads were mapped with STAR (v2.7.3a), if applicable reads were selected with samtools (v1.10) and transcripts were quantified with rsem (v1.3.3)(Dobin et al. 2013; H. Li et al. 2009; B. Li & Dewey 2011). Differential gene expression analysis was performed using the DESeq2 R package (Love et al. 2014). Downstream analysis and figure making were performed also in R.

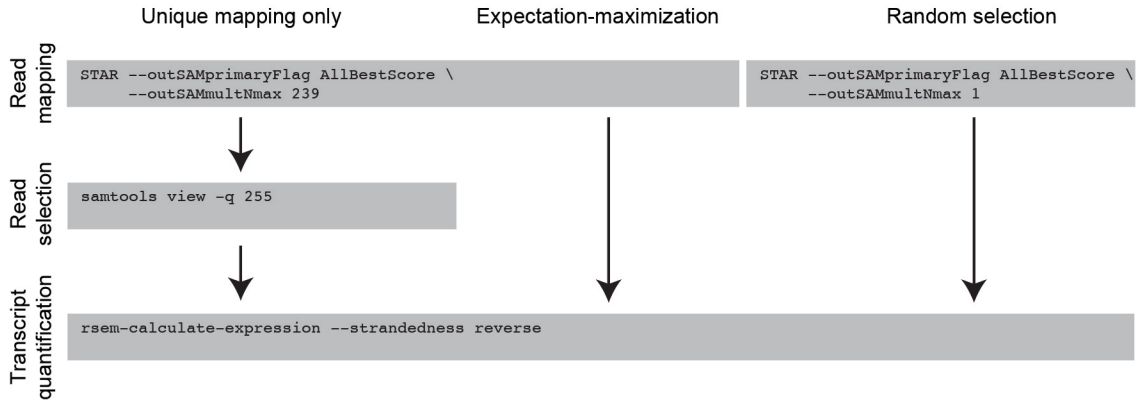


Figure 2.2 Processing of NGS reads for RNA-seq analysis

Schematic overview of NGS read processing for RNA-seq analysis. Software was used at default settings with the exceptions of the settings shown here.

2.13.6 Analysis of published ChIP-seq data set

The *in silico* DNA sequence analysis was performed using Python. The genomic sequence used in this analysis is the S. pombe genomic sequence complemented with the pNSU70 as described above.

2.13.7 Analysis of genome-wide genetic screen data

After straightening the pictures using the ruler function of Adobe Photoshop CS5.1 (Adobe, version 12.1 x64), the picture were cropped using the crop function of preview (Apple Inc., version 8.1). Colony size on the different plates was determined using the R package Gitter. Outer rings on the plates were corrected for "border effect" (larger colonies, probably due to more available nutrients and more initial cells due to concave shape of media). Border effect correction factor is determined by dividing the median size of the not affected colonies and dividing it by the median of the colonies to be corrected. To apply the border effect correction the colony size of the colonies to be corrected is multiplied by the correction factor. Ratios between growth on selective media and growth on non-selective media are then calculated and normalized by dividing by the media ratio of all colonies on the plate. Results from individual plates are combined and finally a table is made combining all screen data (Figure 2.3). This final table is used to calculate average and standard error of the mean, which are plotted.

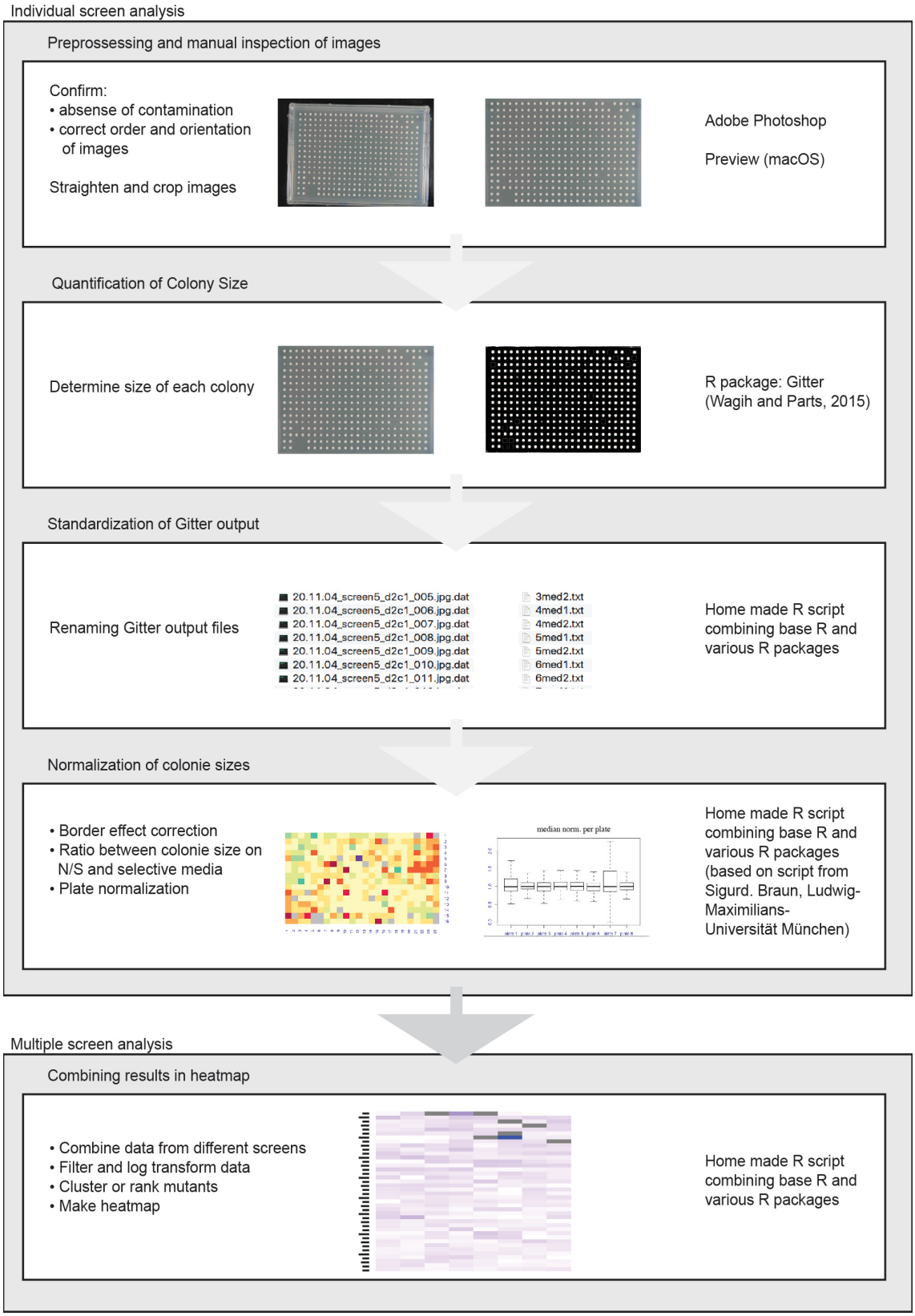


Figure 2.3 Genomic screen analysis pipeline

See text for details

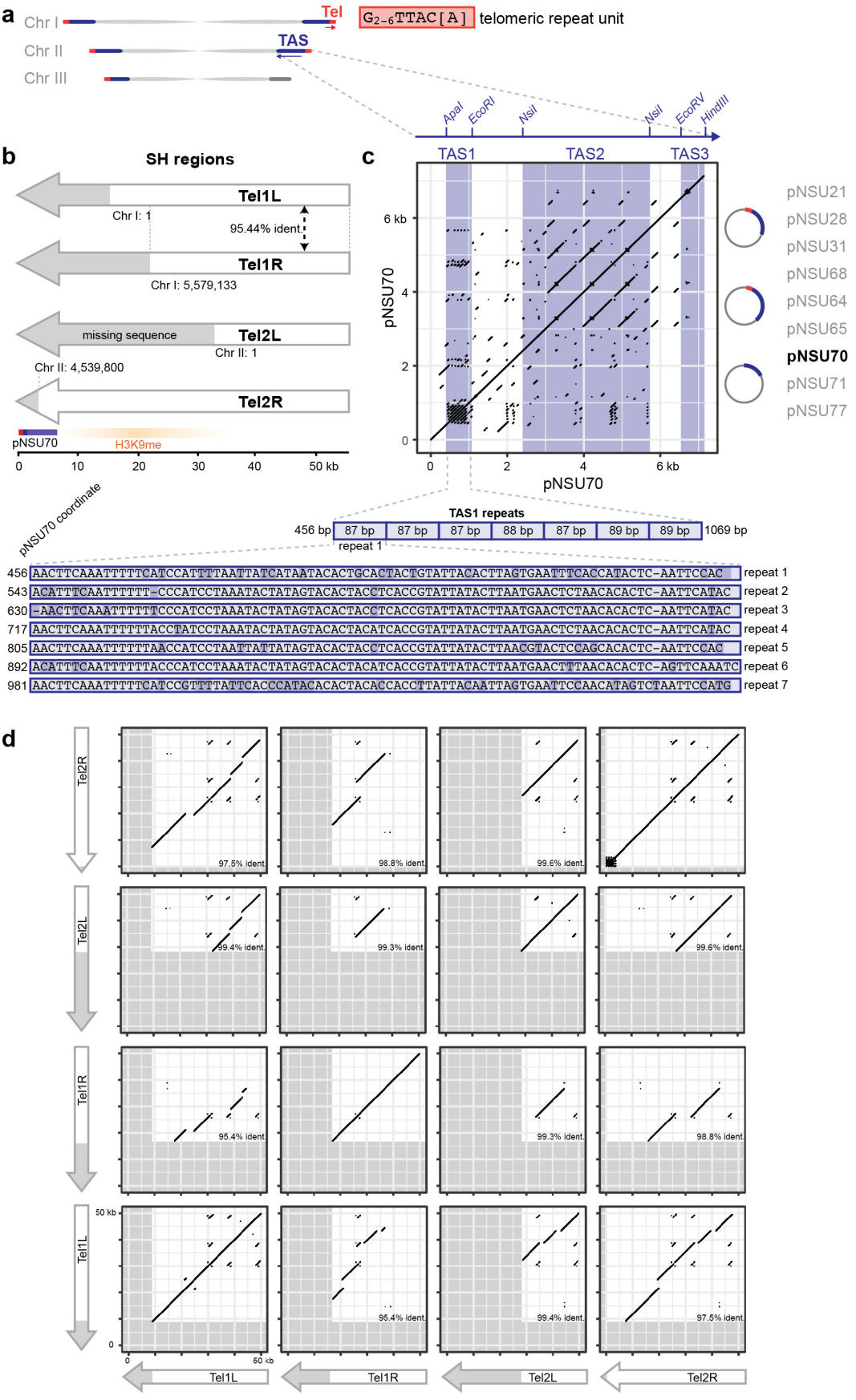
3. Results

3.1 Subtelomeric regulation

3.1.1 The subtelomeres of *S. pombe* contain tandem repeats

While genome sequencing allowed almost complete assembly of the *S. pombe* genome, it did not result in full assembly of the telomere and subtelomere sequences (Wood et al. 2002). In an early attempt to resolve the subtelomeric sequences several fragments were isolated and cloned into plasmids (Sugawara 1988). Among these plasmids is the pNSU70 vector, which contains the complete TAS region of the right arm of chromosome II (Figure 3.1a-c). The TAS regions of *S. pombe* are described as repetitive but the exact nature of the repeats is not fully understood (Sugawara 1988). To gain insights in the nature of the repeats, I aligned the TAS region from on the pNSU70 vector with itself using a dot matrix to identify the repeat regions. I identified two prominent tandem repeat regions. The first tandem repeat region is found in TAS1 with seven blocks of ~87 bp imperfectly repeated sequences (Figure 3.1c). This repetitive region was later also identified by others, albeit they identified 8 blocks of ~87 bp in the same region (Oizumi et al. 2021). The second tandem repeat region was found in TAS2, where three larger blocks of ~500 bp repetitive sequences were identified (Figure 3.1c). These dot matrixes provide insights into the exact nature and positions of the repeats in the TAS regions. It should be noted however, that the alignment presented here displays the TAS region of the right arm of chromosome II. Given that the repeats are not perfect, which is a sign that the TAS region is prone to mutations, it is possible that the TAS landscape of the other chromosome arms is highly similar but not identical to the alignment presented here.

Besides the tandem repeats of the TAS regions it is known that the subtelomeric arms of chromosome I and II themselves are interspersed repeats (Sugawara 1988). To further investigate the homology between the arms of chromosome I and II of *S. pombe,* I made


a
Chr I
Chr II
Chr III
Tel
TAS
G2-6TTAC[A] telomeric repeat unit
ApaI
EcoRI
NsiI
NsiI
EcoRV
HindIII
b
SH regions
Tel1L
Chr I: 1
95.44% ident.
Tel1R
Chr I: 5,579,133
missing sequence
Tel2L
Chr II: 1
Chr II: 4,539,800
Tel2R
pNSU70
H3K9me
0 10 20 30 40 50 kb
pNSU70 coordinate
c
TAS1
TAS2
TAS3
6 kb
4
2
0
pNSU70
0 2 4 6 kb
pNSU70
pNSU21
pNSU28
pNSU31
pNSU68
pNSU64
pNSU65
pNSU70
pNSU71
pNSU77
TAS1 repeats
456 bp
87 bp
87 bp
87 bp
88 bp
87 bp
89 bp
89 bp
1069 bp
repeat 1
456 AACTTCAAATTTTTCATCCATTTTAATTATCATAATACACTGCACTACTGTATTACACTTAGTGAATTTCACCATACTC-AATTCCAC repeat 1
543 ACATTTCAATTTTTT-CCCATCCTAAATACTATAGTACACTACCTCACCGTATTATACTTAATGAACTCTAACACACTC-AATTCATAC repeat 2
630 -AACTTCAAATTTTTTCCCATCCTAAATACTATAGTACACTACCTCACCGTATTATACTTAATGAACTCTAACACACTC-AATTCATAC repeat 3
717 AACTTCAAATTTTTTACCTATCCTAAATACTATAGTACACTACATCACCGTATTATACTTAATGAACTCTAACACACTC-AATTCATAC repeat 4
805 AACTTCAAATTTTTTAACCATCCTAATTATTATAGTACACTACCTCACCGTATTATACTTAACGTACTCCAGCACACTC-AATTCCAC repeat 5
892 ACATTTCAATTTTTTACCCATCCTAAATACTATAGTACACTACATCACCGTATTATACTTAATGAACTTTAACACACTC-AGTTCAAATC repeat 6
981 AACTTCAAATTTTTCATCCGTTTTATTCACCCATACACACTACACCACCTTATTACAATTAGTGAATTCCAACATAGTCTAATTCCATG repeat 7
d
Tel2R
Tel2L
Tel1R
Tel1L
97.5% ident.
98.8% ident.
99.6% ident.
99.4% ident.
99.3% ident.
99.6% ident.
95.4% ident.
99.3% ident.
98.8% ident.
95.4% ident.
99.4% ident.
97.5% ident.
50 kb
0
0
50 kb
Tel1L
Tel1R
Tel2L
Tel2R

<

Figure 3.1 Sequence analysis of the TAS and SH regions

a Locations of TAS and telomeric repeats on the three chromosomes of *S. pombe*: note that the presence of the TAS sequence on chromosome III varies between laboratory strains (Baumann & Cech 2000). The highlighted sequence is a telomeric repeat unit (Dehé & Cooper 2010). The small arrows denote the direction of telomeric (red) and TAS (blue) sequences (the telomeric repeat sequence is shown in the 5′–3′ direction toward the telomeric end). **b** Schematic alignment of the SH regions of chromosome I and II. Tel1R and Tel2R are shown as reverse complement. The white regions indicate sequences present in the current *S. pombe* genome sequence (www.pombase.org; (Wood et al. 2002)), whereas gray regions are missing. Scale bar indicates the distance in kb away from the telomere. Overlap with a subtelomeric sequence retrieved from the cosmid pNSU 70 (Sugawara 1988) and distribution of heterochromatin (H3K9me) based on Tashiro et al. 2017 are shown. Percentage identity (ident.) between Tel1L and Tel1R is indicated, whereas identity between other subtelomeres is shown in d. Percentage identity is calculated using blastn suit-2 sequences. **c** Left panel shows a dot-matrix of the TAS regions illustrating repetitive elements. The TAS sequences are derived from pNSU70, which shows substantial overlap with Tel2R. The matrix was generated using the dotplot package available at https://github.com/evolvedmicrobe/dotplot, with following settings: window size = 87 and number of matches = 58. Shown below is a sequence alignment of seven consecutive 86–89 bp tandem repeat units of TAS1 (starting coordinates refer to pNSU70). Right panel lists cosmids from Sugawara (1988) b and c expand on previous work published in (Chaudari & Huberman 2012; Tashiro et al. 2017). **d** Dot matrix of the SH regions of chromosome I and II shows that the SH regions are homologous and also contain repetitive regions. Gray shaded areas are not present in the *S. pombe* genome sequence (as in b). Setting used: window size = 20; number of matches = 20. Figure and legend taken from (van Emden & Braun 2019).

dot matrixes by aligning the most telomere proximal 50 kb of both chromosomes.

It should be noted that large parts of these sequences are not available in the standard genome assembly of *S. pombe*, and that the missing regions were therefore omitted in this analysis (Figure 3.1b). The analysis revealed that the telomere proximal 50 kb of both arms of chromosome I and II are largely similar (Figure 3.1d). However, as seen in the TAS region, the homology is not perfect, as the arms do contain insertions and deletions as previously reported (Chaudari & Huberman 2012). Since subtelomeric sequences are similar, recombination may take place resulting in gain of heterogeneity of the subtelomeric arms. Overall the degenerate repeats and homology of the subtelomeric arms, in-

cluding the TAS regions, suggests that the subtelomeres are a dynamic chromosomal region that due to recombination can undergo constant change.

3.1.2 TAS regions have low nucleosome occupancy

Studies from Dr. Marta Forn in our group revealed the TAS regions have low levels of H3K9me2 compared to other large heterochromatin domains (van Emden et al. 2019). The underlying reason for this phenomenon remained unclear until Marta Forn found that the reduced H3K9me2 levels at TAS correlate with low nucleosome occupancy in the TAS region (van Emden et al. 2019). This suggests that the low nucleosome occupancy is the underlying cause of the reduced H3K9me2 in the TAS regions. However, so far, we had only studied nucleosome occupancy at the TAS regions using histone H3 chromatin immunoprecipitation followed by qPCR (ChIP-qPCR). Although highly quantitative, this method provides limited insights into the TAS region as a whole. To gain additional insights into the nucleosome occupancy of the most telomere-proximal 20 kb of the right

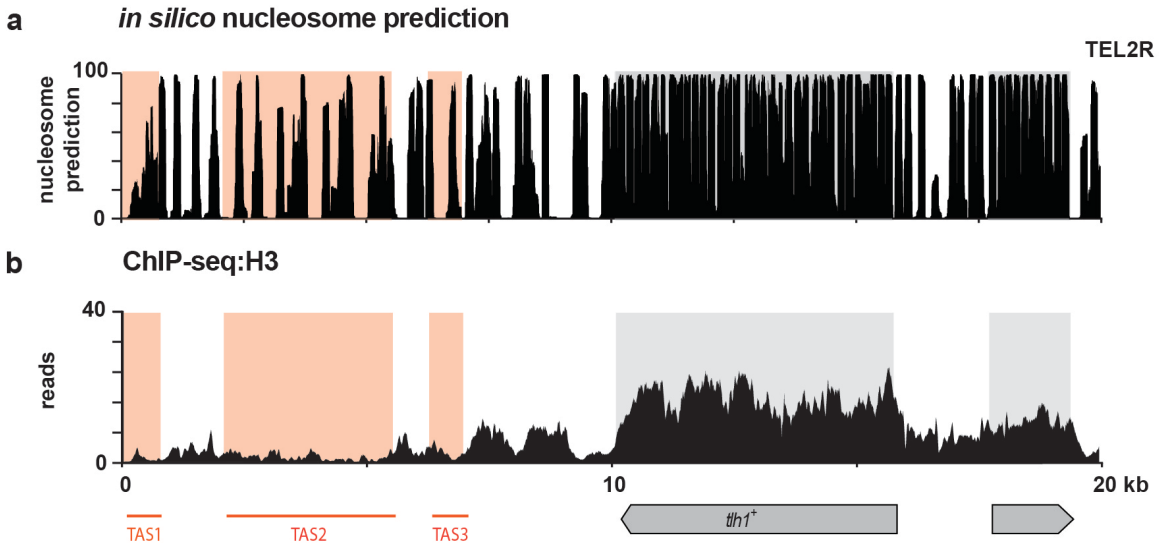


Figure 3.2 Nucleosome abundance is low throughout the telomere-proximal chromosomal region

a Nucleosome prediction for TEL2R using prediction algorithm (Xi et al. 2010). **b** ChIP-seq reads mapped to TEL2R (shown in reverse orientation for consistency). Figure and legend taken from (van Emden et al. 2019).

arm of chromosome II, I used a nucleosome positioning prediction algorithm to calculate the specific location of several nucleosomes within the TAS region (Xi et al. 2010). Since the published *S. pombe* genome sequence does not entail the most telomere proximal sequence it was supplemented with the TAS sequence from the pNSU70 plasmid (Sugawara 1988; Wood et al. 2002). The nucleosome occupancy of the TAS region as a whole is predicted to be much lower compared to neighboring genes (Figure 3.2a). This observation is in agreement with ChIP-qPCR analysis performed by Dr. Marta Forn (van Emden et al. 2019).

To confirm the subtelomere-wide low nucleosome occupancy with experimental data, I performed H3 ChIP-seq (data analysis performed by Cornelia Brönner, group of Prof. Mario Halic, Ludwig-Maximilians-Universität München). The ChIP-seq reads are mapped against the genomic sequence of *S. pombe,* complemented with the TAS sequence of the pSU70 plasmid. This complemented version of the genome thus contains the sequence of the TAS region only once. However, in reality the genome contains four TAS regions distributed among the chromosomal ends of chromosome I and II (Figure 3.1a-b). Therefore, the nucleosome occupancy at the TAS region determined by ChIP-seq cannot be compared with the nucleosome occupancy at euchromatin or other heterochromatic regions in the ChIP-seq experiment. However, nucleosome levels at the TAS region can be compared with *tlh1*$^{+}$, the most telomere proximal gene, as this gene is present at each subtelomere of chromosome I and II. Corroborating previous ChIP-qPCR data, the nucleosome H3 levels in the entire TAS region are low but rise with increasing distance from the telomere and correlate with the low level of H3K9me in the TAS region observed by Marta Forn (Figure 3.2b). Indeed, suggesting that the low level of H3K9me is caused by low level of nucleosomes in the TAS region.

3.1.3 Low nucleosome occupancy is an intrinsic property of the TAS regions

To elucidate if the low nucleosome occupancy is a property of the TAS regions or rather a consequence of its proximity to the telomere, I inserted a 790 bp fragment, containing a large part of the DNA sequence of TAS1, into an ectopic locus at the euchromatic *leu1*$^+$ gene (scheme Figure 3.3a). To exclusively measure the histone H3 levels at this ectopic TAS site, I performed the ChIP experiment in a strain background where all endogenous TAS regions have been deleted (Tashiro et al. 2017). Remarkably, I observed very similar histone H3 occupancies between the ectopic TAS and endogenous TAS (Figure 3.3a). This suggests that the TAS sequence autonomously establishes low H3 occupancy.

To further rule out effects from the telomere proximity, I performed a reciprocal experiment. I measured histone H3 occupancy at euchromatic marker genes (*ura4*$^+$ and *his3*$^+$) inserted between the telomere and TAS (Nimmo et al. 1998). Whereas the TAS has low nucleosome occupancy, the *ura4*$^+$ and *his3*$^+$ reporter genes retain their high, euchromatic H3 levels (Figure 3.3b). Finally, I measured the histone H3 occupancy of a reporter gene inserted into a mini-chromosome in which the subtelomeric region has been deleted (Nimmo et al. 1994). Here, the reporter gene is flanked by a telomere and euchromatic DNA (scheme Figure 3.3b). Again, histone H3 occupancy at the reporter gene was not reduced, ruling out that the TAS region has low nucleosome occupancy due to the proximity of the telomere (Figure 3.3b). Together, these results imply that the low nucleosome occupancy is a feature of the TAS region itself.

3.1.4 Low A-T content correlates low nucleosome occupancy

The DNA sequence of TAS regions is known to be A-T rich and as such is predicted to be refractory to nucleosome binding (Creamer et al. 2014). Therefore, I hypothesized that the high A-T content is the driver of the observed low nucleosome occupancy. To get better understanding of the A-T distribution throughout the TAS regions, I analyzed the

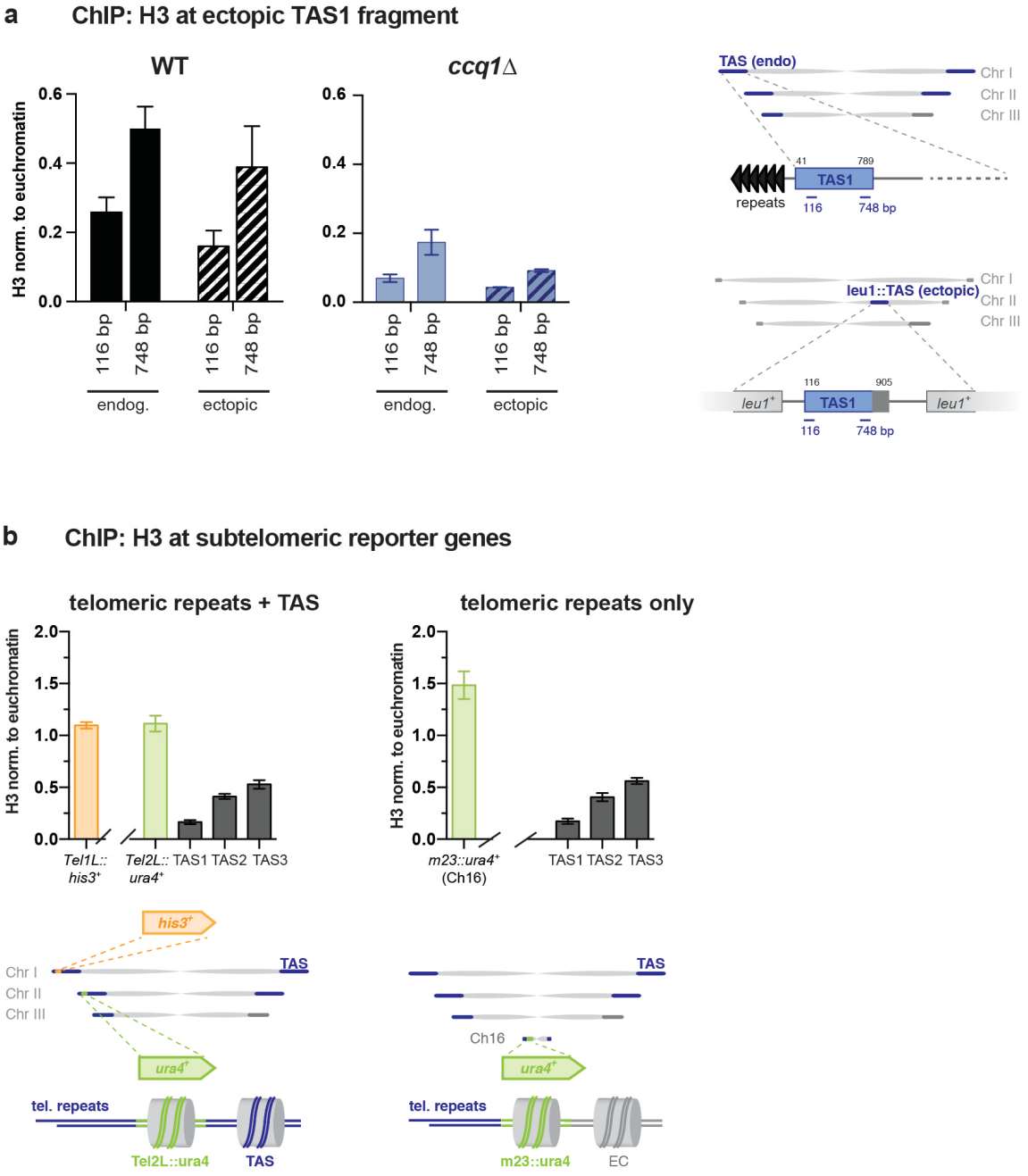


Figure 3.3 Ectopic insertion of TAS fragment is sufficient to cause nucleosome instability

a ChIP-qPCR analysis of H3 at endogenous TAS and a fragment spanning the TAS region from 115 to 905 bp (relative to telomeric repeat) inserted into the leu1+ locus (see scheme). The TAS fragment was inserted into a strain that lacks endogenous TAS (see text)(Tashiro et al. 2017). Shown are ChIP analyses for WT (left) and *ccq1*Δ (right; note different the scale of the y-axis)(n = 9–10 independent experiments except for ectopic TAS in *ccq1*Δ strain where n = 3). **b** ChIP-qPCR analysis of H3 at reporter genes (ura4 + and his3 +) at various chromosomal locations (see schemes) in WT cells. TAS1, TAS2, and TAS3 correspond to position 116, 2851, and 6291 bp (relative to telomeric repeats), respectively (n = 3 independent experiments). Data are represented as mean ± SEM. Figure and legend taken from (van Emden et al. 2019).

A-T content of the same 20kb region that was used for the nucleosome prediction algo rithm and mapping of the ChIP-seq reads (see Figure 3.2). The average A-T content in *S. pombe* is ~64%, but the majority of the TAS region, especially TAS2 and TAS3, has a higher A-T content, in some places more than 90% (Figure 3.4a). Such high A-T content is not observed in the *tlh1*$^{+}$ gene or in more telomere-distal regions.

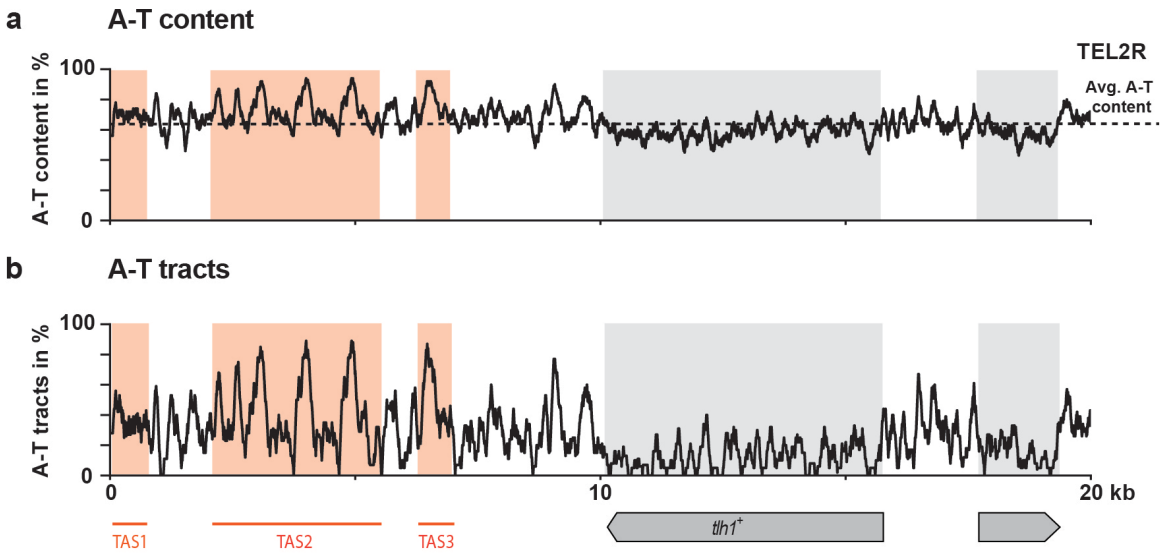


Figure 3.4 TAS regions are characterized by high A-T content

a *In silico* analyses of A-T content of TEL2R DNA. Black line shows A-T content. Dotted line represents average A-T content in *S. pombe* (Wood et al. 2002). Red and gray shaded areas show TAS regions and subtelomeric genes, respectively. **b** *In silico* analyses of A-T tracts in TEL2R DNA. Shown is the percentage of poly[A-T] tracts (defined as a sequence of 5 or more nucleotides consisting only of A or T). Red and gray shading as in a. Figure and legend taken from (van Emden et al. 2019).

A-T tracts (defined as a sequence of 5 or more nucleotides consisting only of A or T) are known to poorly incorporate into nucleosome due to their rigidity (Struhl 1985; Nelson et al. 1987; J. D. Anderson & Widom 2001). I found that the percentage of bases in A-T tracts is higher in the TAS region than for example in the *tlh1*$^{+}$gene (Figure 3.4b). The locations of high A-T content and A-T tracts strongly correlate with low nucleosome occupancy found the H3 ChIP-seq experiment (Figure 3.2b). This implies that the high A-T

tracts in the TAS regions could be the underlying cause of the low nucleosome occupancy in the TAS regions.

3.1.5 RSC does not decrease subtelomeric nucleosome levels

However, additional mechanisms could also lead to instability of nucleosomes at the TAS region, for example the Chromatin Structure Remodeling (RSC) complex. This complex promotes nucleosome eviction (J. F. Garcia et al. 2010; Creamer et al. 2014). Therefore, I studied the histone H3 occupancy in a temperature-sensitive RSC mutant, as the deletion of the RSC catalytic core (*snf21*Δ) is inviable (Yamada et al. 2008). In the temperature sensitive mutant strain, at restrictive temperature (34 °C), the nucleosome level at TAS regions rises marginally compared to the wild-type strain, but it does not reach levels observed for euchromatin (Figure 3.5). Rather, the nucleosome level at the euchromatic locus that was probed in the RSC mutant seems to rise in a similar fashion as the nucleosome level at the TAS. This suggests that the increase in nucleosome occupancy observed upon perturbation of the RSC complex is not specific for the TAS region but rather a genome-wide phenomenon.

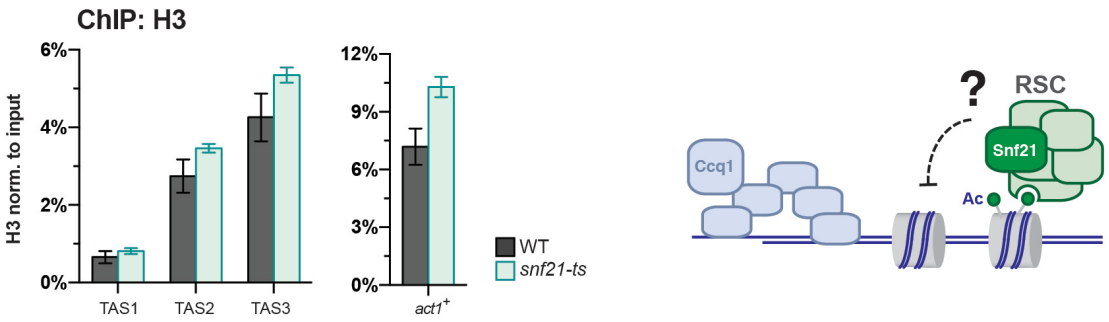


Figure 3.5 Low nucleosome occupancy at TAS is not caused by the RSC

ChIP-qPCR analysis of H3 in WT and snf21-ts cells (see scheme)(n = 3 independent experiments each derived from 2 to 3 parallel ChIP samples). Data are represented as mean ± SEM. Figure and legend taken from (van Emden et al. 2019).

3.1.6 Shelterin antagonizes TAS-mediated nucleosome instability

The shelterin subunit Taz1 binds not only telomeric sequences, but also the subtelomeric region, up to at least 9 kb from the telomeric repeats (Cooper et al. 1997; Sadaie et al. 2003). It is conceivable that the low nucleosome occupancy at the TAS region is a result of competition for binding of the TAS sequence between nucleosomes and shelterin. Therefore, I probed the nucleosome occupancy of the TAS regions in light of two shelterin mutants. Instead of leading to an increase in nucleosome levels in the TAS regions, the levels of histone H3 further decreased in the mutants of DNA binding shelterin component Taz1 and telomerase recruiting shelterin component Ccq1 (Figure 3.6a). This suggests that the shelterin complex, rather than nucleosome instability, causes nucleosome stability in TAS regions.

To investigate if the TAS DNA sequence is sufficient to mediate nucleosome stability and regulation by shelterin in the absence of telomeres, I studied the role of Ccq1 in nucleosome stability at the ectopic TAS region inserted in the *leu1*$^{+}$ locus. The ectopic TAS region autonomously establishes low nucleosome levels, very similar to the endogenous TAS regions. Likewise, the ectopic and endogenous TAS regions display a remarkably similar reduction in nucleosome occupancy in the *ccq1*Δ mutant (Figure 3.3a). Although direct binding of shelterin to the ectopic TAS has not been studied here, these data sug-

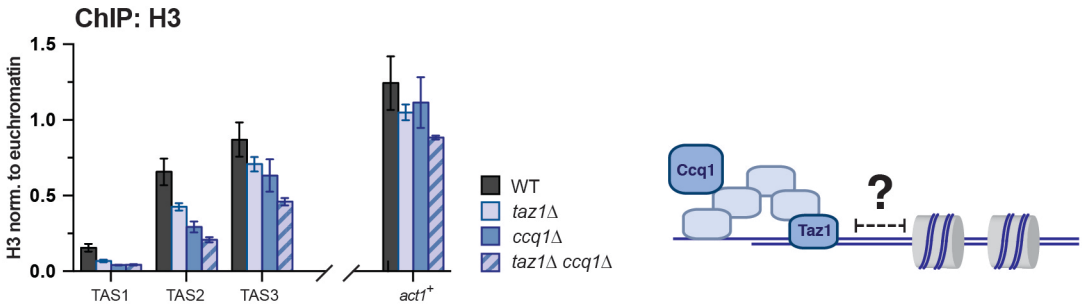


Figure 3.6 Shelterin stabilizes nucleosomes at the TAS

a ChIP-qPCR analysis of H3 in WT, taz1Δ and ccq1Δ and corresponding double mutants (see scheme)(n = 3 independent experiments). Data are represented as mean ± SEM.

gest that nucleosome instability in the TAS region is an inherent property of its DNA sequence and that shelterin counteracts this instability.

To determine how shelterin mediates nucleosome stability Dr. Marta Forn identified physical interactors of shelterin using a proteomics approach. This mass spectrometry experiment retrieved Clr3, component of SHREC, as the main interactor of shelterin. This validated the proteomics approach, as shelterin is known to bind this heterochromatic effector complex (Sugiyama et al. 2007). Additionally, all components of the CLRC methyltransferase complex were identified as interactors of shelterin. Around the same time this discovery was made in our lab it was also published in Wang et al., (2016) and Zofall et al., (2016).

Clr4, a component of the CLRC complex, is the sole H3K9 methyltransferase in *S. pombe* (Nakayama et al. 2001). H3K9me is recognized by members of the HP1 family, Swi6 and Chp2, that mediate spreading of heterochromatin (Noma K et al. 2001). However, Marta Forn showed that the HP1 proteins do not seem to play a role in the maintenance of nucleosomes at the subtelomeres of *S. pombe*.

We next hypothesized that heterochromatin complexes SHREC and CLRC could affect nucleosome stability at the TAS region as they are recruited there by shelterin. Indeed, Marta Forn found that mutants of SHREC and CLRC have reduced nucleosome occupancy of the TAS regions, similar to shelterin mutants. Furthermore, mutants of $ccq1^{+}$ combined with mutants of either SHREC or CLRC show an epistatic phenotype with respect to nucleosome occupancy at the TAS region. Interestingly, we found that deletion of shelterin, SHREC or CLRC does not cause instability of nucleosomes at the most telomere proximal gene, $tlh1^{+}$. Thus, we concluded that shelterin maintains nucleosomes as the TAS regions through the recruitment of heterochromatin complexes SHREC and CLRC (Figure 3.7).

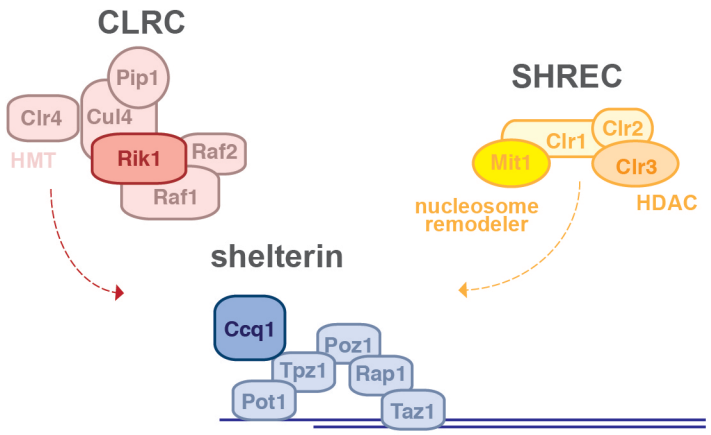


Figure 3.7 The shelterin complex and its binding partners in *S. pombe*

Overview of the proteins that constitute the shelterin complex (blue), CLRC (red) and SHREC (yellow). The shelterin complex binds DNA (blue lines) at single stranded 3' telomeric overhang through Pot1 and double stranded DNA through Taz1. CLRC and SHREC are recruited (arrows) to the (sub)telomere by the shelterin complex through Ccq1. Figure adapted from (van Emden & Braun 2019).

3.1.7 Catalytic activity of Mit1 is required for nucleosome stability at the TAS regions

Since we found that the shelterin complex recruits SHREC to the TAS regions I wondered if it is the remodeler activity of Snf2-like, ATP-dependent nucleosome remodeler Mit1, component of the SHREC complex, that is responsible for nucleosome stability. To investigate this, I probed the nucleosome occupancy of the TAS regions in wild-type, *mit1*Δ and a *mit1-K587A* catalytically dead point mutant cells (see scheme Figure 3.8). The point mutant showed similar reduced nucleosome occupancy as the *mit1*Δ (scheme Figure 3.8). This finding shows that the catalytic activity of Mit1 is required to maintain nucleosomes at the TAS region.

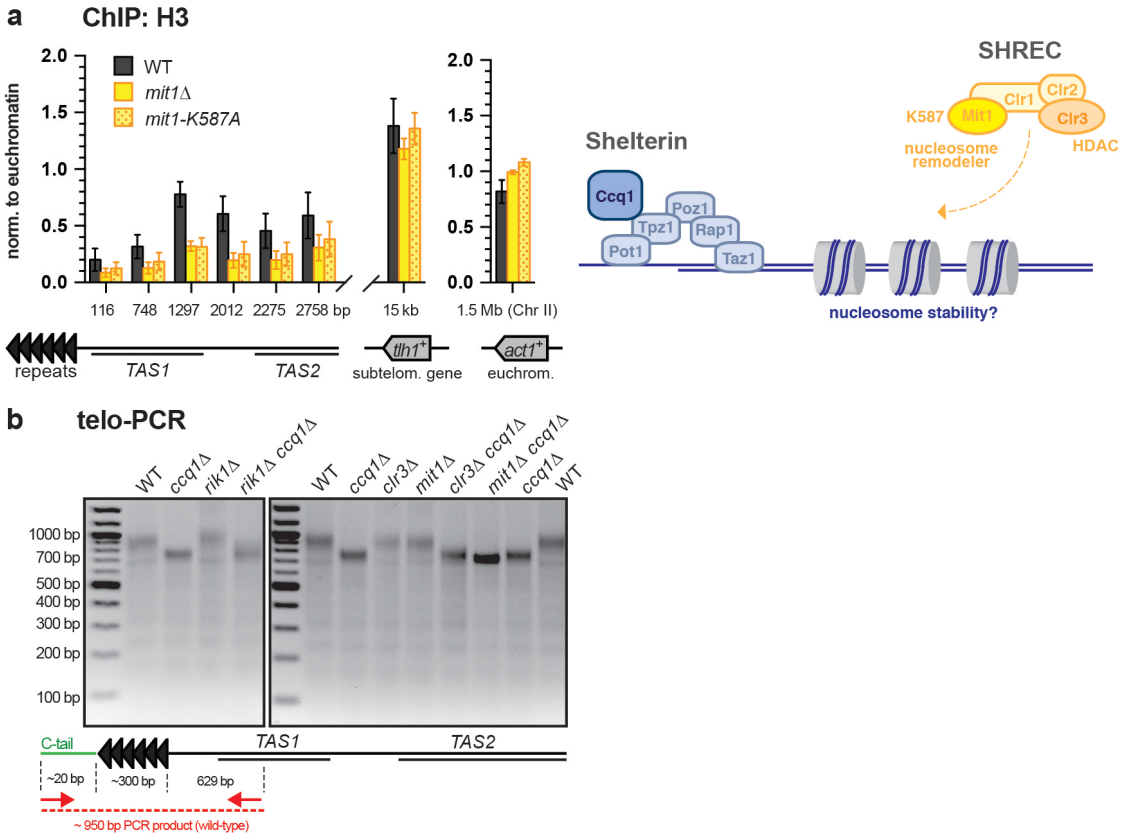


Figure 3.8 Catalytic activity of Mit1 is required for nucleosome stability at the TAS regions

a ChIP-qPCR analyses of H3 levels in WT and mutants as indicated (n = 3 independent experiments). Data are represented as mean ± SEM. Figure and legend taken from (van Emden et al. 2019). **b** Telomere length analysis of different mutants using telo-PCR, which amplifies the entire repeats and 629 bp of the adjacent TAS1 region (Moravec et al. 2016). PCR products are analyzed on a 0.8% agarose gel. Note that telomeres are shortened in early *ccq1*Δ (~200 nt) but still retain telomeric repeats. Figure and legend taken from (van Emden et al. 2019).

To ensure that the lowered nucleosome occupancy observed in the TAS regions in shelterin, SHREC and CLRC mutants is not caused by shortening of telomeres and subtelomeres, I determined the telomeric length in a wild-type strain and different shelterin mutants. To do so I applied a method called Telomere PCR (Telo-PCR) that relies on poly(C) tailing of genomic DNA. Next a PCR reaction is performed using a subtelomeric primer and a primer that anneals on the border between the telomere and the poly(C) tail. The size of the PCR fragment can be used to determine the length of the telomere, as the distance between the subtelomeric primer and telomere is known (Moravec et al. 2016).

Using telo-PCR I showed that deletion of *ccq1*$^+$ results in shortening of telomeres by approximately 200 bp. Since the telomeric length in *S. pombe* is approximately 300 bp, the subtelomeres remain intact upon deletion of *ccq1*$^+$ (Figure 3.6b). Double mutants of *ccq1*$^+$ and SHREC and *ccq1*$^+$ and CLRC also have shortened telomeres but single SHREC and CLRC mutants do not have shorted telomeres, suggesting that the shortening in the double mutants is caused by mutation of *ccq1*$^+$ alone. Thus, the decreased nucleosome occupancy in the TAS region observed in shelterin single and double mutants cannot be explained by shortening of the subtelomere.

3.1.8 Shelterin, CLRC and SHREC repress subtelomeric transcription

Nucleosome stability is an important factor in transcription (Workman 2006). The TAS region encodes TERRA, a lncRNA that is transcribed and plays a role in the regulation of telomere length)(Figure 3.9a). Therefore, I employed RT-qPCR to study the role of shelterin, SHREC and CLRC in the regulation of TERRA transcription.

*mit1*Δ and *ccq1*Δ cells show comparable transcriptional upregulation of TERRA (Figure 3.9b). Additionally, the double mutant shows an epistatic genetic interaction between *mit1*$^+$ and *ccq1*$^+$. Deletion of *clr3*$^+$ and *rik1*$^+$ lead to a slightly stronger TERRA derepression than the *ccq1*$^+$ deletion. Furthermore, the *clr3*$^+$-*ccq1*$^+$ double deletions show partial additive effects. Thus, at the TAS regions, shelterin establishes silencing of TERRA largely through SHREC and CLRC; however, *clr3*$^+$ and *rik1*$^+$ may also act partially independently of *ccq1*$^+$. Interestingly at further distance from the telomere, at the *tlh1*$^+$ locus, shelterin and SHREC and CLRC show synthetic phenotypes, suggesting that shelterin, in this region works through different mechanism (Figure 3.9c). As expected, more telomere distal, at different heterochromatin domains, for example the centromeres, shelterin does not play a role in regulation of transcription (Figure 3.9d).

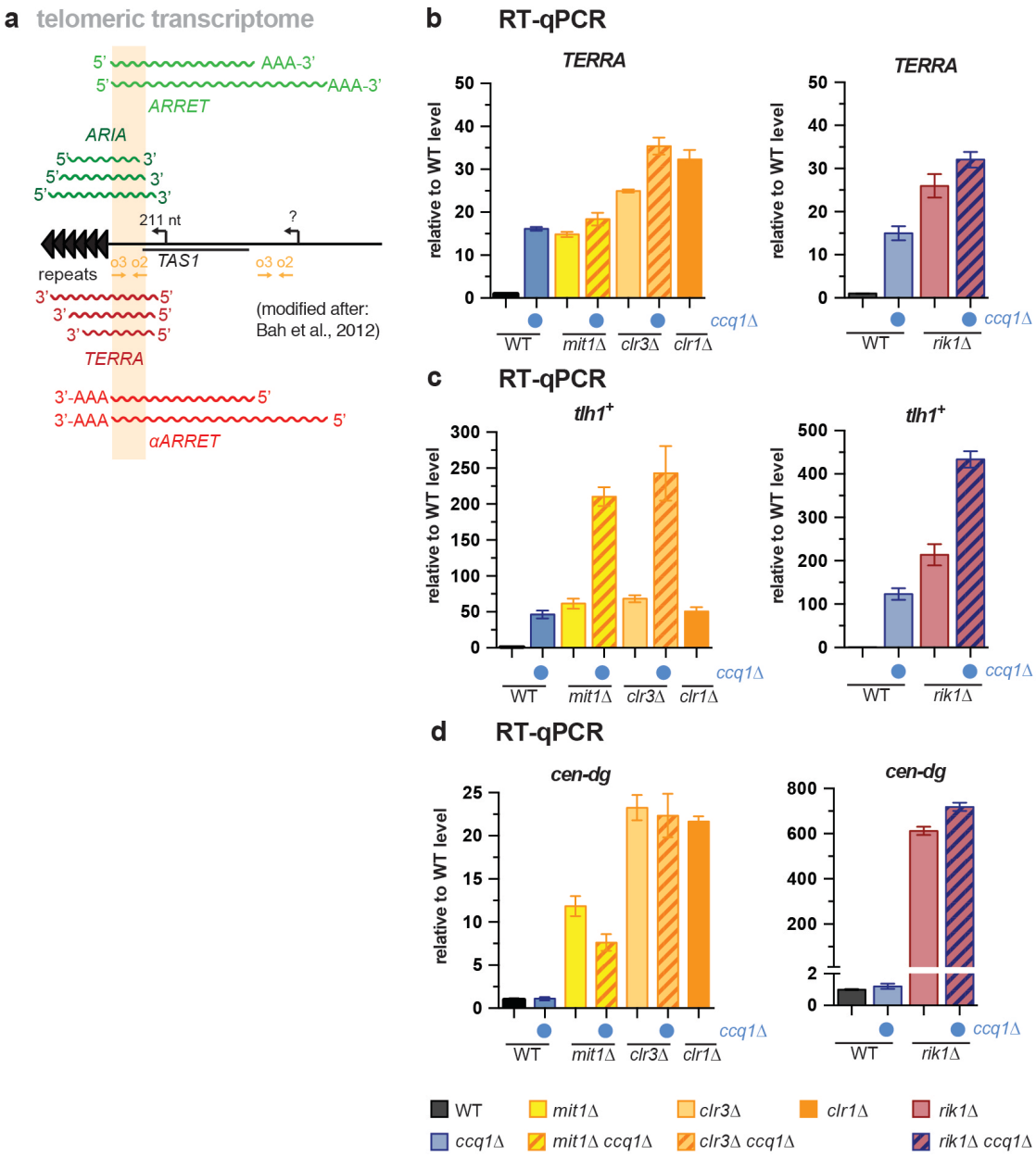


Figure 3.9 Nucleosome stability at subtelomeres is linked to heterochromatic silencing

a Scheme of telomeric transcripts (modified after (Bah et al. 2012)). Whereas ARRET and αARRET RNAs have a poly-A tail, only a small percentage of TERRA and ARIA transcripts are poly-adenylated. The amplicon (primers o2 + o3) used for RT–qPCR and ChIP-qPCR analysis anneals to all telomeric transcripts without discriminating strand specificity or shorter species that lack transcribed parts from the telomeric repeats. For simplicity, these transcripts are referred to as "TERRA". An identical sequence of the o2/o3 amplicon is present in a telomere-distal region, but it is unknown whether this region also contains transcription start sites. **b-d** RT–qPCR analysis of transcript levels of TERRA (b), tlh1+ (c) and cen-dg (d) in WT strain and mutants as indicated (double mutants with ccq1Δ are indicated by blue dot)(n=3). Data are represented as mean ± SEM from three independent experiments and shown relative to WT level. Figure and legend taken from (van Emden et al. 2019).

3.1.9 Shelterin represses subtelomeric recombination but not through nucleosome maintenance

It is known that the TAS regions are prone to genomic rearrangements in *ccq1*Δ strains (Tomita & Cooper 2008). This leads to the question if the homologous sequences of the TAS regions contribute to the genomic rearrangements between chromosomal arms. However, studying interchromosomal rearrangements between subtelomeres is difficult due to their high level of similarity. Therefore, I monitored recombination through a non-subtelomeric euchromatic gene, used as an individual barcode, inserted in between a telomere and a TAS region on the left arm of chromosome II. Recombination between chromosomal arms will lead to a loss or gain of the barcode-marked subtelomere. Therefore, I monitored the copy number of the barcode over 6 days in wild-type and *ccq1*Δ cells (Figure 3.10a). In the wild-type strain the genomic copy number of the reporter gene remained constant, however, in the *ccq1*Δ the reporter gene was highly unstable (Figure 3.10b). As expected the genomic copy number of the TAS appeared to remain stable throughout this experiment. The genomic instability of the reporter gene is in line with earlier reports that in the absence of Ccq1 the subtelomeres undergo increased recombination and are highly unstable (Tomita & Cooper 2008).

A previous study revealed frequent genomic rearrangements but did not examine what drives this instability or which sequences are involved (Tomita & Cooper 2008). To address this question, I studied whether genomic rearrangements depend on the TAS region. To this end, I studied the genomic copy number of a barcode inserted next to a telomere on a mini-chromosome without TAS regions. This mini-chromosome contains telomeres but lacks the entire subtelomeric region (Nimmo et al. 1994). Using the same approach as described above, I found that the barcode on the mini-chromosome was more stable compared to the barcode inserted between a telomere and a TAS region (Figure 3.10b-c). While this finding implies that the TAS regions facilitate recombination between the sub-

telomeres, it does not show whether the TAS regions are required, as other homologous parts of the subtelomere might also be able to facilitate recombination.

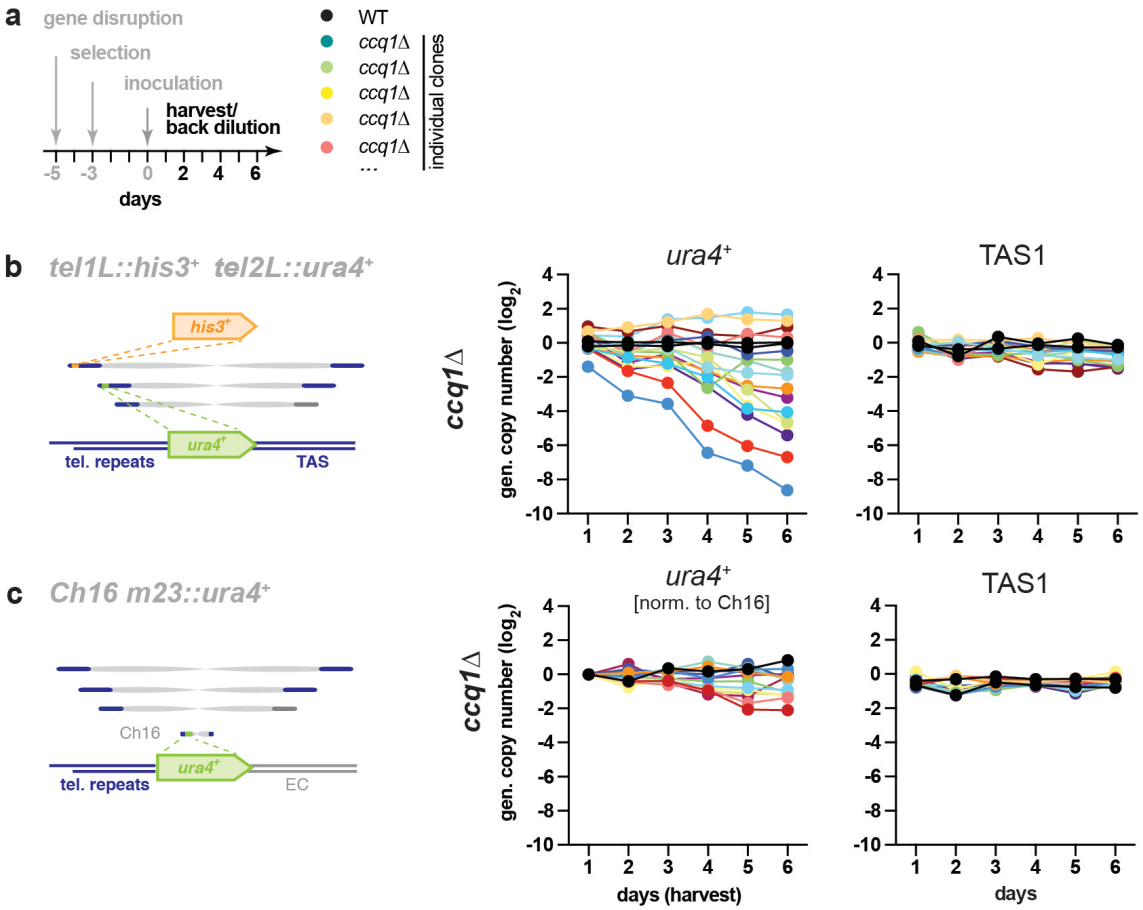


Figure 3.10 Subtelomeric DNA promotes recombination in the absence of Ccq1

a Set up of the genomic stability assay. Cultures from individual WT and freshly generated knockout clones were pre-grown on selective media for several days and inoculated at day 0 to grow in liquid media with regular back-dilution (every 24 h, approximately 7 generations). Samples were taken at indicated harvest times, and relative copy numbers of genomic regions were assessed by qPCR (normalization against intrachromosomal loci). **b** Genomic copy number of reporter gene inserted between the telomeres and the TAS region. Black and rainbow color lines indicate WT strains and individual *ccq1*$^{+}$ (n= 16) deletion mutants (clones), respectively. **c** as in b, but the reporter gene is inserted on he minichromosome Ch16 m23::ura4+ that lacks TAS regions. Figure and legend taken from (van Emden et al. 2019).

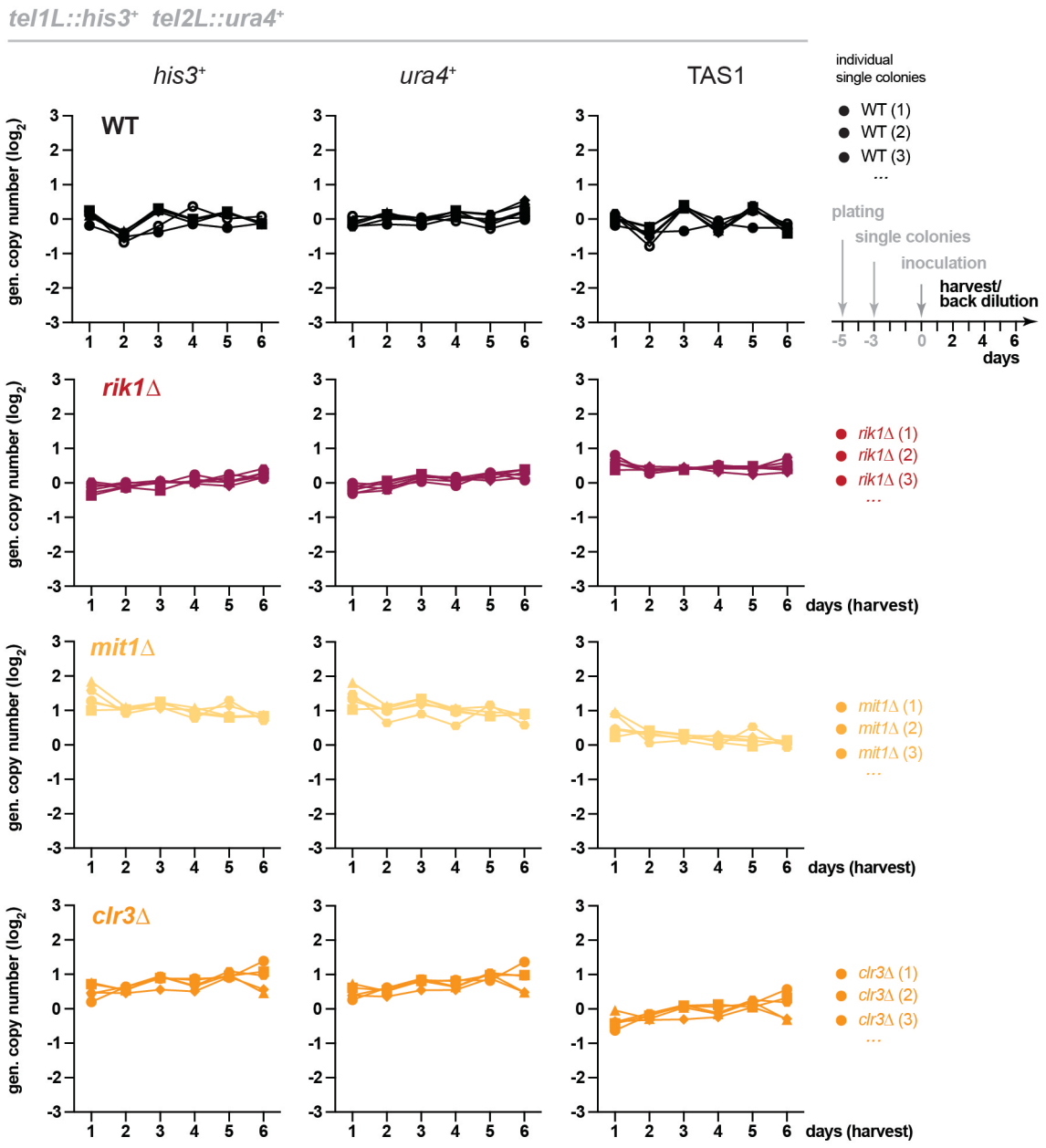


Figure 3.11 Subtelomeres do not undergo recombination in the absence of CLRC and SHREC

Genomic copy number of his3 +, ura4 +, and TAS1 in indicated strains harboring the reporter genes tel1L::his3 + and tel2L::ura4 + cells. Cultures from individual WT and knockout clones (WT, n = 6; rik1Δ, n = 6; mit1Δ, n = 5; clr3Δ, n = 5) were inoculated at day 0 to grow in liquid media with regular back-dilution (every 24 h, approximately 7 generations). Samples were taken at indicated harvest times, and relative copy numbers of genomic regions were assessed by qPCR (normalization against intrachromosomal loci). Figure and legend taken from (van Emden et al. 2019).

Ccq1 is critical to maintain nucleosomes stability in the TAS region, this raises the question whether nucleosome occupancy is linked to genomic stability. To address this ques-

tion, I performed genome stability experiments as described above in strains lacking SHREC and CLRC. The reporter gene was found to be equally stable in the wild-type and mutants strains indicating that Ccq1 maintains subtelomeric stability through a different mechanism than maintaining nucleosome stability (Figure 3.11).

3.1.10 Summary of subtelomeric regulation in *S. pombe*

In summary, compared to subtelomeric genes and euchromatin, the repetitive TAS regions are characterized by low nucleosome occupancy. From my bioinformatical analysis of the TAS sequence, I conclude that it seems likely that the nucleosome instability is caused by the high A-T content of the TAS regions. The shelterin complex through the recruitment of SHREC and CLRC prevents additional loss of nucleosomes. Through analysis of transcription of TERRA, a subtelomere encoded lncRNA, I demonstrate that the maintenance of nucleosomes by shelterin, SHREC and CLRC represses subtelomeric transcription. However, by performing subtelomere stability assays I show that shelterin does not require SHREC and CLRC to maintain genomic stability of the subtelomeres, suggesting telomere stability is regulated via different mechanism then nucleosome stability.

3.2 Regulation of LTRs

3.2.1 Dissecting the LTR regulatory pathways

Besides the repetitive subtelomeric region, the genome of *S. pombe* also contains more then 200 interspersed repeats in the form of full-length LTR retrotransposons and solo LTRs (Bowen et al. 2003). Given their potential to retrotranspose and recombine, these repetitive elements form a threat to genome stability. Therefore *S. pombe* employs various mechanisms to silence retrotransposons and LTRs. Some of these mechanisms have

been well studied, whereas other mechanisms remain elusive. Furthermore, most characterized LTR silencing factors are predominantly studied in the context of full-length retrotransposons, thus the regulation of solo LTRs is not well understood. Finally, a systematic analysis of factors controlling LTR silencing and how they cooperate is lacking. As a first step, I measured LTR expression by RT-qPCR in mutants of various LTR regulators. In order to find pathways of LTR regulation the resulting LTR expression patterns were used to correlate the LTRs and their regulators. This approach is based on the rational that mutants of factors that cooperate likely have similar LTR expression patterns.

3.2.1.1 Primer design for repetitive regions

The genome of *S. pombe* contains 13 full-length Tf2 retrotransposons and more than 200 solo LTRs, of which 28 and 35 of Tf1 and Tf2 origin, respectively. The repetitiveness of these LTRs poses a challenge for the design of qPCR primer pairs that target only a single, specific LTR. However, LTR repeats of *S. pombe* are not perfect. This allowed me to design specific qPCR primer pairs for a large number of LTRs. I designed the primer pairs used in this study using primer-BLAST, an online tool hosted by the National Center for Biotechnology Information (Ye et al. 2012). This tool combines primer design using the primer3 software and basic local alignment search tool (BLAST) alignments of the resulting primers to confirm that only a single product can be amplified in a qPCR experiment (Koressaar & Remm 2007; Altschul et al. 1990). The selected LTRs are evenly dispersed over the genome, with no bias for chromosome or chromosomal location (Figure 3.12a, black vertical lines).

To confirm that the designed primer pairs are specific to a single LTR, I analyzed the resulting qPCR products by agarose gel electrophoresis and analyzed their melting

curves. All qPCR products showed a single, clear band of expected size when analyzed by agarose gel electrophoresis (Figure 3.12b).

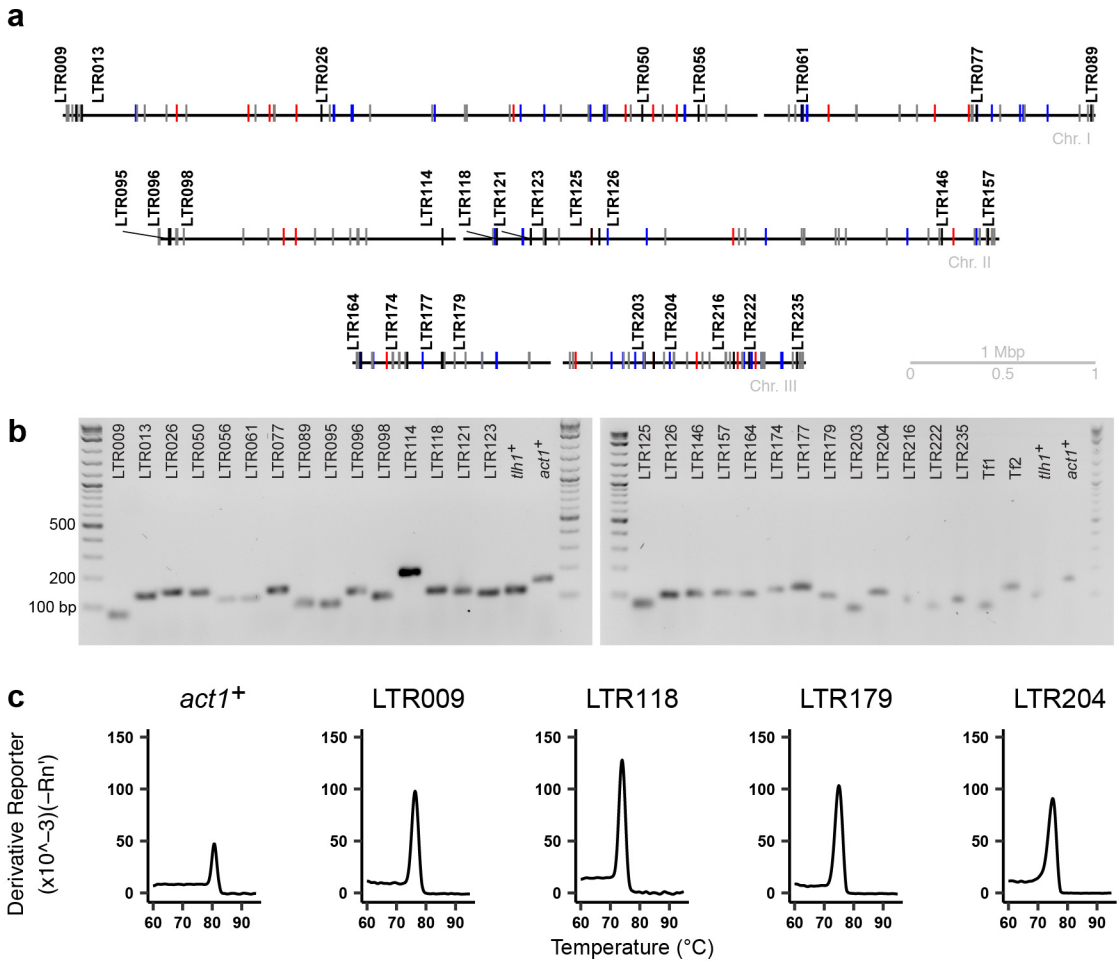


Figure 3.12 Quality control of qPCR primers that target LTRs

a Localization of the LTRs in the genome of *S. pombe*. Three horizontal black lines represent the three chromosomes of *S. pombe,* gaps in the black lines represent the centromeres. Individual LTRs probed in this study are depicted as black vertical lines (several individual LTRs later used in the study are labeled). Tf1 LTRs are shown as red vertical lines, Tf2 LTRs are shown as blue vertical lines and all other LTRs are shown as grey vertical lines. **b** LTR qPCR products analyzed by agarose gel electrophoresis. Primer pairs as indicated on top of the lane. Outermost lanes contain a DNA ladder with indicated sizes. **c** Melting curves of qPCR products of a euchromatic control (*act1*$^{+}$) and four LTRs (representative for the four LTR clusters identified in 3.2.1.3).

However, the detection of a single band does not exclude the possibility that multiple LTRs were amplified. One can imagine that qPCR products originate from different genomic loci but have similar length. This holds true especially for the repetitive LTRs.

Therefore, in addition, melt curves of the qPCR products were studied. If multiple LTRs with different DNA sequences are amplified in a qPCR reaction, the melting curve of the qPCR product will have 2 peaks or a single peak with a shoulder. I observed single peaks in this analysis but the peaks are slightly wider than the non-repetitive control (*act1*$^{+}$). This result is consistent with the assumption that the majority of the LTR qPCR product originates from a single LTR (Figure 3.12c).

Some LTRs show such high sequence similarity that specific primer design is not possible, for example the LTRs of Tf1 and Tf2 origin. To include the Tf1 and Tf2 LTRs in this analysis, primer pairs were designed that are specific for each group but not individual LTRs, recognizing all (28 and 35, respectively) Tf1 and Tf2 LTRs (Figure 3.12a, red and blue vertical lines respectively). It should be noted that the primer pair targeting Tf2 targets both solo Tf2 LTRs and the LTRs of Tf2 full-length retrotransposons. Overall, the primer design resulted in 28 primer pairs that recognize a single solo LTR and 2 primer pairs that recognize the Tf1 and Tf2 LTRs.

3.2.1.2 Solo LTRs are regulated by retrotransposon regulating factors

Several factors have been identified as regulators of full-length retrotransposons in *S. pombe*. These factors include a CENP-B specific silencing pathway, histone chaperones, targeting factors of the nuclear exosome and chromatin remodelers (Cam et al. 2008; H. E. Anderson et al. 2009; Yamanaka et al. 2013; Steglich et al. 2015). However, it is not known if all these factors also regulate solo LTRs. Therefore, I probed LTR expression by RT-qPCR in 9 different mutants of known full-length retrotransposons regulators to investigate if solo LTRs are also regulated by these factors. Each LTR has a unique profile of expression across the different mutants (Figure 3.13a and Appendix A). This presence of mutant-specific changes in expression results in individual expression profiles for

each LTR, further supporting the notion that the selected primer pairs recognize distinct targets. More importantly, this shows that solo LTR expression is regulated by the factors that also regulate full-length retrotransposon expression.

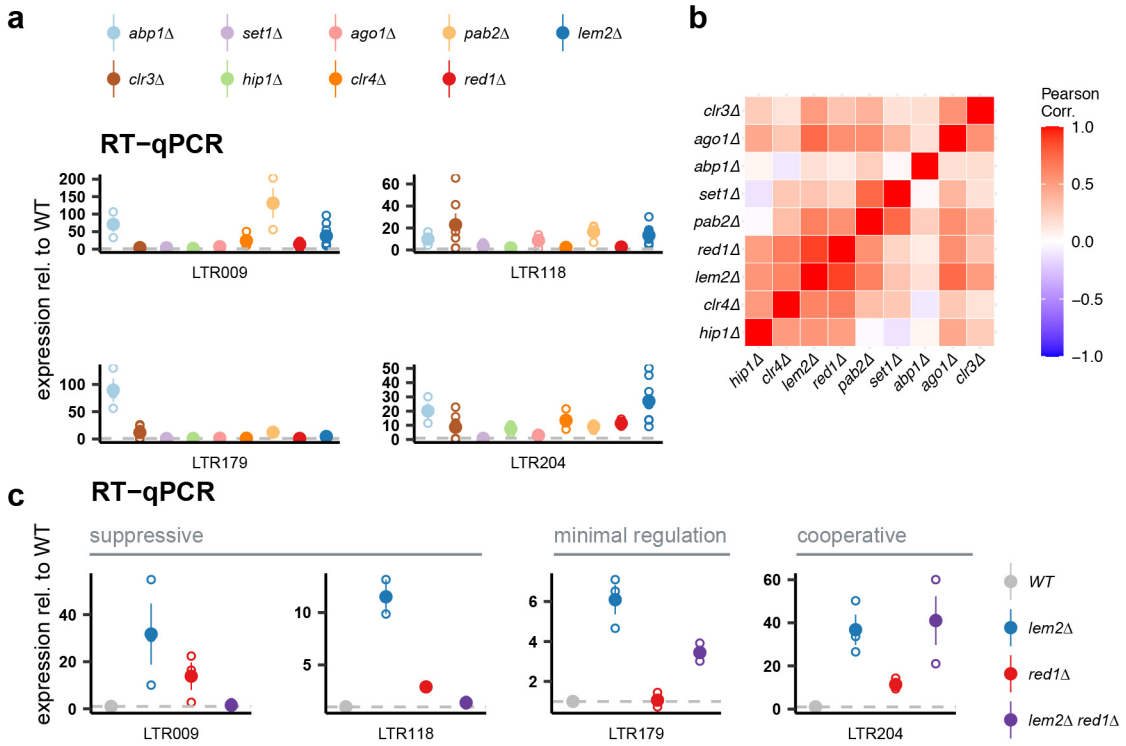


Figure 3.13 Correlation plot reveals functional LTR regulating pathways

a Expression of four LTRs (representative for the four LTR clusters identified in 3.2.1.3) in mutants of known factors involved in LTR regulation (n = 9 for *lem2*Δ and n = 3 for all other mutants). Open circles indicate individual observations, closed circle indicates the average and the vertical line indicates the standard error of the mean. **b** Correlation plot of mutants of known LTR regulators based on LTR expression in the respective mutant. **c** Expression of four LTRs (representative for the four LTR clusters identified in 3.2.1.3) in wild-type (WT), *lem2*Δ, *red1*Δ and the respective double mutant (n =3). Grey text indicated the genetic interaction of *lem2*$^{+}$ and *red1*$^{+}$ with respect to the regulation of these LTRs. Open circles indicate individual observations, closed circle indicates the average and the vertical line indicates the standard error of the mean. This work was performed with support of Agnisrota Mazumder, a fellow PhD student in the lab.

To study whether these factors can be attributed to specific regulatory pathways, I performed unsupervised hierarchical clustering based on LTR expression in the mutants of LTR regulators (Figure 3.13b). This revealed varying levels of correlation between the between the mutants of LTR regulators (Figure 3.13a and Appendix A). Overall, strong

correlation was observed between *lem2*Δ, *red1*Δ*, clr4*Δ *and hip1*Δ cells (Figure 3.13b). The strongest correlation was observed between mutants of inner nuclear membrane protein Lem2 and exosome targeting factor Red1. On the other hand, *hip1*Δ has the lowest correlation with *set1*Δ. Additionally, *abp1*Δ and *clr3*Δ do not correlate strongly, which is surprising, as CENP-B proteins and Clr3 have been reported to act in the same pathway with respect to full-length retrotransposon silencing (Cam et al. 2008).

To further study the genetic interaction between *lem2*$^+$ and *red1*$^+$, I studied the LTR expression in a *lem2*Δ*red1*Δ double mutant strain. This epistasis analysis revealed three main groups of LTRs (Figure 3.13c). The first group (exemplified by LTR009 and LTR118) showed upregulation both in the *lem2*Δ and *red1*Δ strains but suppression in the *lem2*Δ*red1*Δ double mutant strain. The second group of LTRs constitutes LTRs that are predominantly regulated by Abp1 but not by Lem2 or Red1 (exemplified by LTR179). In the third group (exemplified by LTR204), LTRs are similarly upregulated in the *lem2*Δ and *lem2*Δ*red1*Δ double mutant cells. In *red1*Δ cells, LTRs are also upregulated but to a lesser extent than the single and double *lem2*Δ mutant cells. This indicates that Lem2 and Red1 cooperate in the silencing of this subset of LTRs. However, it is not clear if these regulators work through the same mechanism.

3.2.1.3 LTRs clusters into different groups based on different regulatory pathways

Since I observed high correlation of LTR expression between *lem2*Δ, *red1*Δ*, clr4*Δ and *hip1*Δ cells, I decided to further analyze this LTR expression data. I asked if the LTRs themselves also cluster in a correlation matrix based on their expression in the different mutants. Interestingly, this revealed four main clusters of similarly regulated LTRs (Figure 3.14 and Appendix A). Closer examination revealed that cluster 1 is predominantly regulated by exosome targeting factor Pab2. Abp1 contributes to cluster 3, which in-

cludes the Tf2 LTRs. This is line with a previous study that reported that Abp1 regulates Tf2 full-length retrotransposons (Cam et al. 2008). Lem2 and a combination of other factors control LTRs of cluster 4. For cluster 2 it is not immediately clear which factor is predominantly responsible for its regulation. Instead this cluster seems to be regulated by a combination of Abp1, Clr3, Lem2 and Pab2. In the subsequent studies and figures, each cluster is represented by a selected LTR from this group. LTR009, LTR118, LTR179 and LTR204 are the representative LTRs for cluster 1,2,3 and 4 respectively (Figure 3.14 and Appendix A).

The identification of distinct LTR expression clusters raises the question of what determines whether or not a factor regulates an LTR. One can hypothesize that a specific sequence motif in an LTR is responsible for its recognition by an LTR regulator. For example, Abp1 is known to bind two consensus motifs (A1 and A2) in LTRs. Furthermore, several transcripts targeted to the nuclear exosome by Red1 contain a many copies of a motif known as the determinant of selective removal (DSR) sequence, which is recognized by the YTH protein Mmi1)(Table 15). Therefore, I used the FIMO algorithm of the MEME suite to search for these motifs in the solo LTRs that I studied by RT-qPCR (Grant et al. 2011). As expected, the A1 and A2 binding motifs were mostly found in Cluster 3, the Abp1-dependent cluster (Table 15). However, the A1 motif is also frequently found in Cluster 1 and Cluster 2. The DSR does occur in some of the LTRs tested here, however only once or twice per LTR, in contrast to exosome target that are marked with many copies of this sequence (Chen et al. 2011; Yamashita et al. 2012)(Table 15).

Table 15 Occurrence of known degradation motifs in LTRs

Motif	Cluster 1 (6 LTRs)	Cluster 2 (5 LTRs)	Cluster 3 (7 LTRs)	Cluster 4 (10 LTRs)
A1 **TAATATAATA**	50%	60%	71%	30%
A2 **TAATACAATA**	33%	40%	57%	20%
DSR **U(U/C)AAAC**	50%	40%	29%	60%

To discover novel motifs in the LTR clusters I used the MEME algorithm of the MEME suite (Bailey & Elkan 1994). However, for all 4 LTR cluster, this analysis retrieved no cluster-specific motifs; rather, the software identified motifs that are specific to all LTRs.

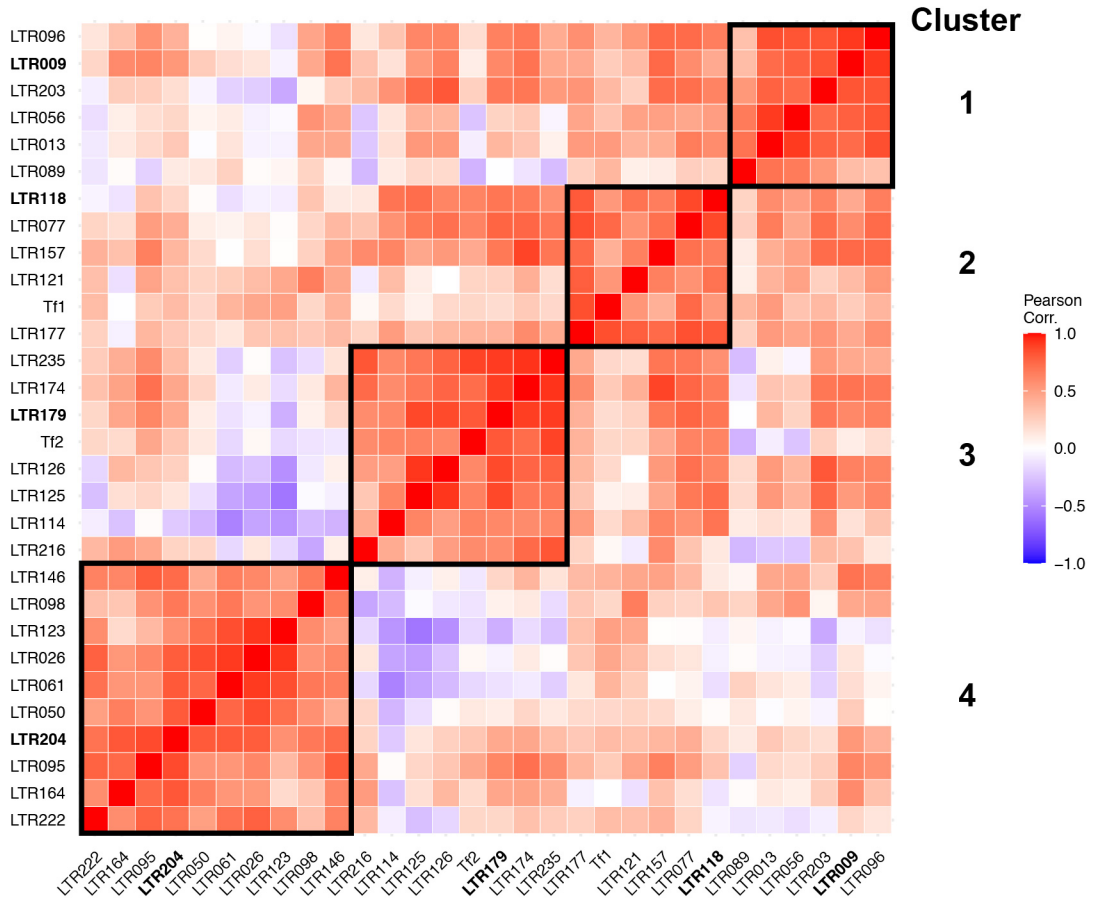


Figure 3.14 Correlation of LTR expression reveals 4 LTR clusters.

LTRs are clustered based on their expression in the known mutants of LTR silencing factors. Four clusters are indicated with black squares. Note that also outside the clusters there is high correlation between LTRs. The four representative LTRs that are used throughout this thesis are highlighted in bold.

LTRs from different clusters are equally distributed between euchromatic and hetero-chromatic regions and also do not display enrichments near specific genomic elements

like telomeres or centromeres. Thus, no specificity towards a genomic location was found for any of the clusters. Furthermore, while the expression level of LTRs in the wild-type strain varies, it does not show a specific pattern within the LTR groups. Thus, the characteristic that determines regulator specificity for most LTRs remains unclear.

3.2.2 Lem2 regulates LTRs genome-wide but not LTR-neighboring genes

Many of the factors probed in this experiment are known to regulate full-length retrotransposons. The inner nuclear membrane protein Lem2, however, has been reported to affect solo LTRs at subtelomeres (Barrales et al. 2016). My results demonstrate that LTR regulation by Lem2 is not limited to the subtelomeric LTRs (Figure 3.13a and Appendix A). The role of Lem2 in LTR regulation is not well understood and was thus studied here in more detail.

To gain better insight in the regulation of LTRs by Lem2 on a genome-wide scale, I performed an RNA-seq experiment in wild-type and *lem2*Δ cells. LTRs pose a specific challenge for the analysis of RNA-seq data, as algorithms fail to map reads to the repetitive regions. Therefore, the RNA-seq data was analyzed using three different methods. The first method reports only unique alignments and omits reads that map to multiple regions. The second methods reports all possible alignment of a read but uses an expectation-maximization algorithm to assign multi-mapping reads based on unique mapping reads (B. Li & Dewey 2011). In the third method one of the regions to which a multi-mapping read maps is randomly selected and reported (for details see section 2.13.5). Another challenge when studying LTR expression by NGS is the low expression of LTRs. LTRs with zero or very few reads (due to mapping issues or low expression) have been omitted from this analysis and, thus, from the 239 annotated LTRs, the expression of only 31, 53

and 56 LTRs is reported using each of the methods, respectively (Figure 3.15a, black dots).

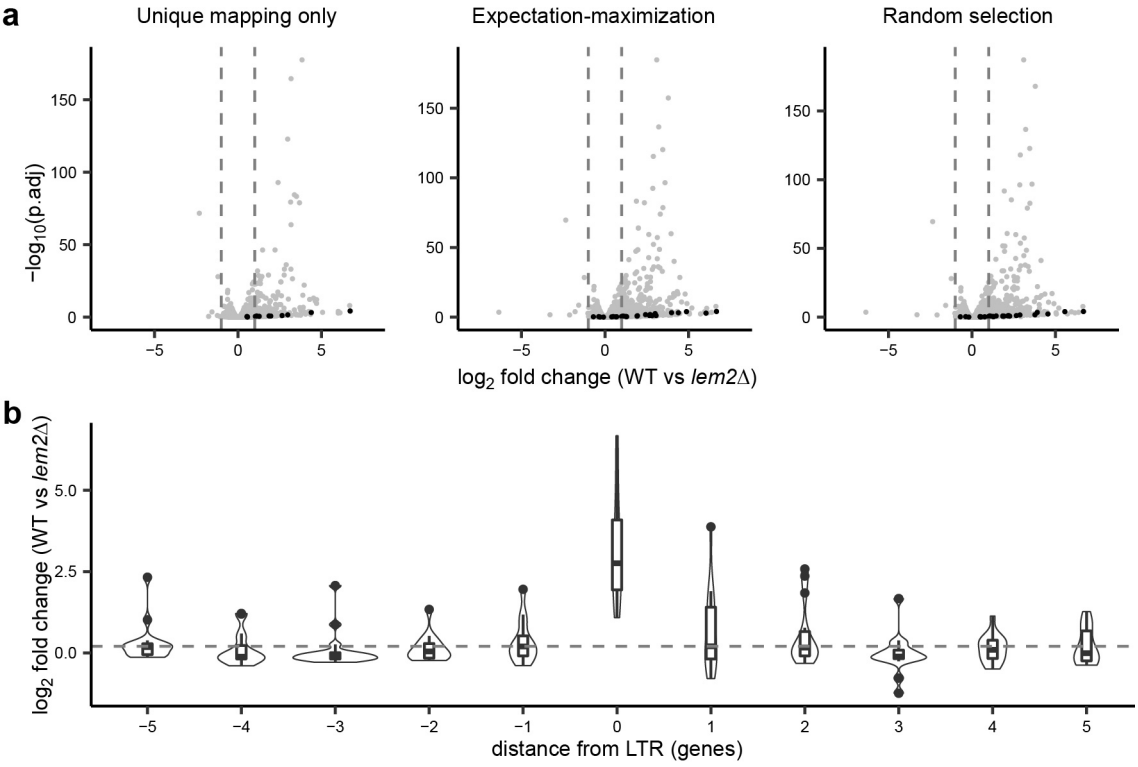


Figure 3.15 Lem2 regulates expression of LTRs but not of LTR-neighboring genes

a Volcano plots of fold change in wild-type (WT) vs *lem2*Δ plotted against adjusted p value (p.adj). Black dots LTRs and all other transcripts are represented by the grey dots. The most upregulated transcripts in the *lem2*Δ cells are ncRNAs and meiotic transcripts, that are further studied in Martín Caballero et al. (2021). The vertical dashed lines mark 2 fold up or down regulated (-1 and 1 on a log2 scale). **b** Violin plots depicting the change in expression of genes neighboring LTRs that are more than 2-fold upregulated in *lem2*Δ cells.

All three mapping methods show that the majority of LTRs is upregulated (>2 fold) in *lem2*Δ cells (58%, 62% and 57% respectively). Although each of the three analysis pipelines outputs a different number of upregulated LTRs, each method reports more than 50% of the LTRs to be at least 2 fold upregulated in *lem2*Δ cells. These LTRs are distributed across the genome, which suggests that Lem2 is a genome-wide regulator of LTR expression, a notion that is consistent with the RT-qPCR results (Figure 3.13a and Appendix A). Expectation maximization algorithms are among the most common ways to

quantify transcripts in RNA-seq experiments and thus for the remainder of this work this method will be used. Again, the DNA sequences of upregulated LTRs were subjected to a *de novo* motif search using the MEME Suite, but this did not reveal any particular motif that differentiates Lem2-regulated LTR from other LTRs.

LTRs serve as promoter regions of retrotransposons or retroviruses and thus can drive gene expression. Additionally, previous publications show that LTRs have promoter activity that leads to transcription of neighboring genes (Sehgal et al. 2007; Cam et al. 2008).

To test if genes in the direct genomic vicinity of LTRs with greater than 2-fold upregulation in *lem2*Δ cells are similarly affected, I studied their relative transcript levels using the same RNA-seq datasets. This revealed very limited increased expression of LTR-neighboring genes, up- and downstream, upon deletion of *lem2*$^{+}$ (Figure 3.15b). Thus, the regulation by Lem2 seems to be very specific for LTRs, nevertheless the mechanism by which Lem2 mediates LTR repression remains unknown.

3.2.3 Hypothesis-driven approach to unravel the LTR silencing mechanism of Lem2

Several hypotheses can be formulated with respect to the mechanism by which LTRs are regulated by Lem2. For example, (1) Lem2 could regulate the expression of another factor that directly regulates LTRs; (2) Lem2 could recruit LTRs to the nuclear periphery through direct binding; (3) Lem2 could be involved in the establishment of heterochromatin; (4) Lem2 could recruit HDACs to the LTRs; or (5) Lem2 could regulate remodelers or histone chaperones. I tested several of the possible mechanisms from this non-exhaustive list.

To test the first hypothesis, if Lem2 regulates the transcription of a known LTR regulator, I examined the most differentially regulated transcripts ($\log_2$ fold change ≤ -1 or $\log_2$ fold

change ≥ in Figure 3.15a) in *lem2*Δ cells. Lem2 seems to function mainly in gene repression as many genes are upregulated in *lem2*Δ cells and only few are down regulated. Together with Lucia Martín Caballero, a fellow PhD student in the lab, I found that ncRNAs and meiotic transcripts are prevalent among the upregulated transcripts (log_2 fold change

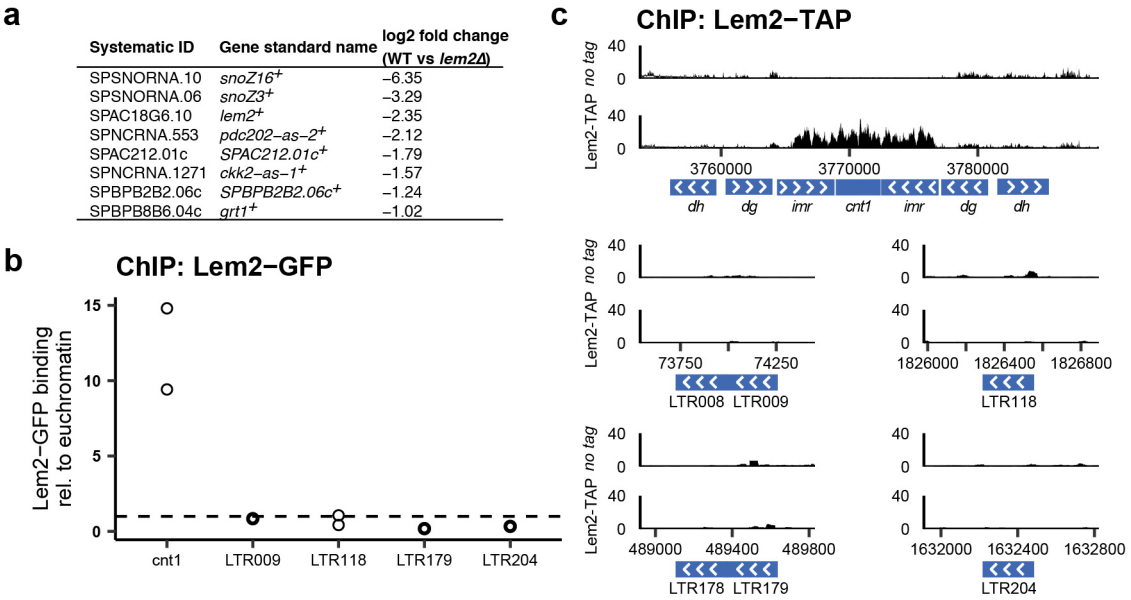


Figure 3.16 ChIP analysis does not reveal direct interaction between Lem2 and LTRs
a Table of the most downregulated transcripts in *lem2*Δ cells. **b** ChIP-qPCR Lem2-GFP binding data (n = 2), open circles indicate individual observations. LTRs are representative for the four LTR clusters identified in 3.2.1.3 **c** ChIP-seq Lem2-TAP binding data. Genome browser screenshots of a no tag control and Lem2-TAP. To allow comparison, tracks are shown on the same scale for centromeric region and all LTRs. Data from Iglesias et al. (2020)

≥ 1 in Figure 3.15a). This analysis, published in Martín Caballero et al. (2021), revealed that only a few genes are down regulated in *lem2*Δ cells. Eight transcripts are downregulated (log_2 fold change ≤ -1 in Figure 3.15a) and are shown in Figure 3.16a. Among those transcripts is *lem2*$^+$, confirming the deletion of gene in the mutant strain. Other transcripts with greatly reduced expression are two small nucleolar RNAs (snoZ16, snoZ3) and antisense transcripts of *pdc202*$^+$ and *ckk2*$^+$, their sense transcripts code for a predicted pyruvate decarboxylase and calcium-calmodulin-dependent protein kinase, respectively. Additionally, two coding genes are down regulated: SPAC2.01c, SPBPB2B2.06c. Final-

ly, in *lem2*Δ cells the gene coding for DNA binding transcription factor Grt1 is down regulated. None of these differential ly expressed transcripts has been published to regulated LTRs.

However, *grt1*$^{+}$ has been reported to be implicated in the silencing of centromere outer repeats, yet its function remains unclear (S. Y. Lee et al. 2020). The role of Grt1in LTR silencing provides an interesting avenue for a follow-up study.

Next, I tested the hypothesis that Lem2 regulates the expression of LTRs by tethering them to the nuclear periphery. Lem2 tethers centromeres via its LEM domain to the nuclear periphery; however, this has been shown to be unrelated to gene repression (Barrales et al. 2016). Nevertheless, I probed whether Lem2 interacts with LTRs by ChIP-qPCR. As expected, I observed no association between the tested LTRs and Lem2, suggesting that Lem2 does not associate with LTRs. Compared to euchromatic control loci, I observe that LTRs are actually depleted. This may suggest that chromatin comprising LTRs is less accessible for the antibody used in the immunoprecipitation. It is tempting to speculate that this may be caused by the nature of the Tf bodies in which retrotransposons cluster together. However, it is not known if solo LTRs are also located in Tf bodies.

Using ChIP-qPCR only a limited number of LTR can be probed for interaction with Lem2. To gain insights on the genome-wide scale, I studied a published Lem2 ChIP-seq dataset (Iglesias et al. 2020). This also revealed no evidence for direct interaction between Lem2 and LTRs except those that are located within the subtelomeric domains that Lem2 is known to interact with. It should be noted that LTR depletion from chromatin associated with Lem2, as seen by ChIP-qPCR, was not seen by genome-wide experiments. This discrepancy could be caused by different sensitivity of the two approaches. In summary, in both the ChIP-qPCR and ChIP-seq data no evidence was found that suggests that Lem2 directly interacts with LTRs.

Previous work from the Braun Lab showed that Lem2 contributes to heterochromatin silencing (Barrales et al. 2016). H3K9me-mediated heterochromatin is involved in the silencing of TEs in many organisms. This leads to the hypothesis that Lem2 silences LTRs through regulation of heterochromatin. In *S. pombe,* H3K9 is methylated by the sole histone methyltransferase Clr4. When testing the genetic relation between *lem2*$^+$ and *clr4*$^+$ with respect to LTR silencing, I observed increased expression in two out of four LTRs I tested in the *clr4*Δ mutant (Figure 3.17a). This finding was unexpected, as H3K9me was not globally detected at LTRs in a genome-wide study (Cam et al. 2005). The four tested LTRs all had increased expression in *lem2*Δ cells. Interestingly, I observed that LTRs in the double mutant of *lem2*$^+$ and c*lr4*$^+$ have higher expression than in either of the single mutants (Figure 3.17a). I hypothesized that deletion of *lem2*$^+$ causes H3K9me accumulation at LTRs, partially compensating the silencing defect of *lem2*Δ; additional deletion of c*lr4*$^+$ could prevent methylation of H3K9 leading to the observed synthetic genetic interaction seen for the double *lem2*Δ c*lr4*Δ *mutant.* Such a synthetic phenotype would be reminiscent of observations made for Tf2 transposons located in so-called HOODs (heterochromatin domains). These Tf2 transposons accumulate H3K9me upon deletion of *rrp6*$^+$, in an RNAi dependent manner, which suggest a synthetic interaction between *rrp6*$^+$ and *clr4*$^+$. However this synthetic interaction has not been studied (Yamanaka et al. 2013).

To test if deletion of *lem2*$^+$ causes H3K9me accumulation at LTRs, I examined H3K9me2 levels in wild-type and *lem2*Δ cells. However, deletion of *lem2*$^+$ did not lead to increased levels of H3K9me2 at LTRs (Figure 3.17b). This makes it less likely that a compensatory mechanism acts in parallel and gets activated in the absence of Lem2 to establish H3K9me. Overall, these data show that heterochromatin does not play a direct role in Lem2-mediated LTR silencing, however an indirect regulation might be in place,

although this is currently not understood.

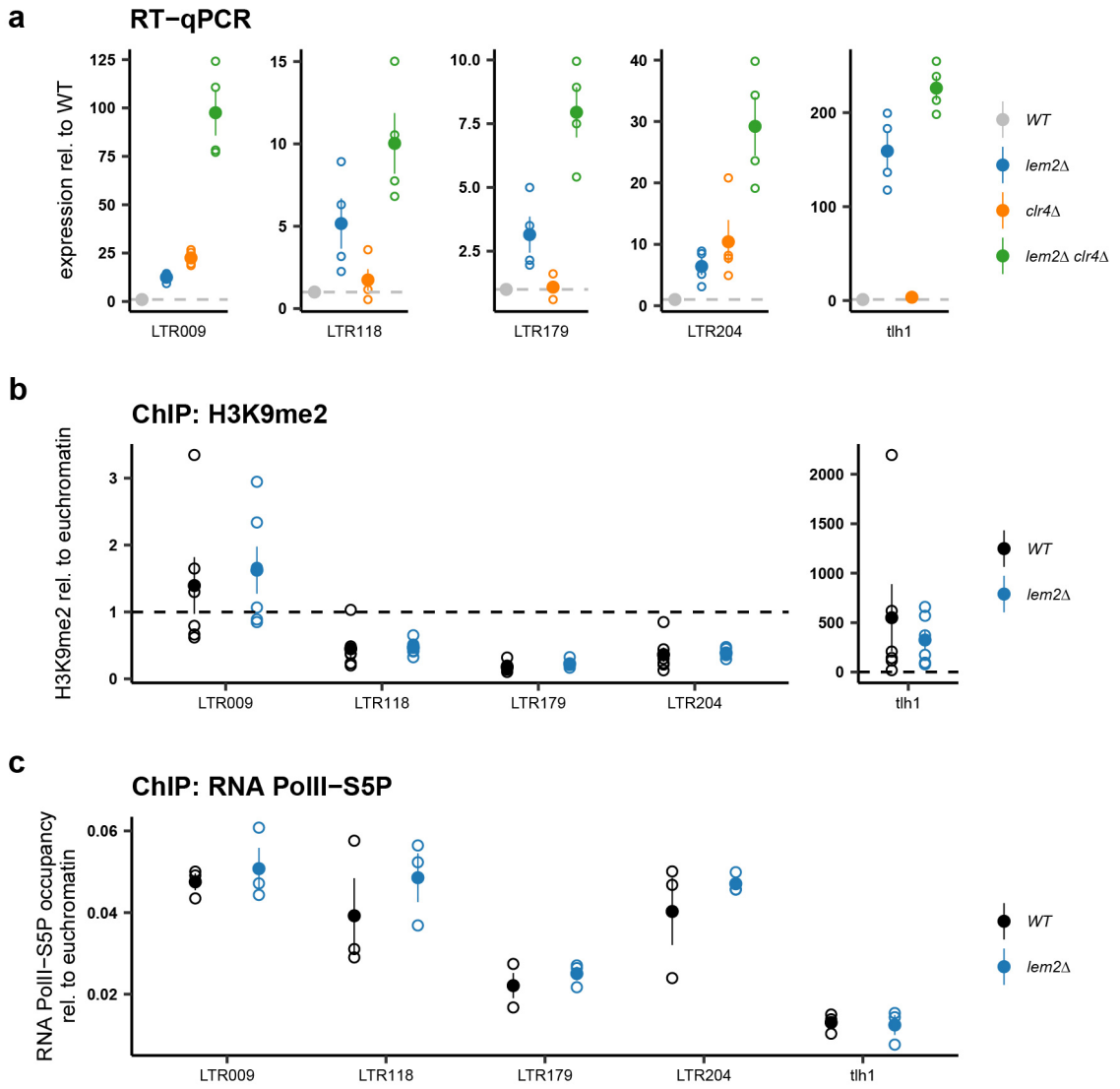


Figure 3.17 Minor increase in occupancy of actively transcribing RNA PolII in *lem2*Δ cells at LTRs

a Expression of four LTRs (representative for the four LTR clusters identified in 3.2.1.3) in wild-type (WT), *lem2*Δ, *clr4*Δ and the respective double mutant (n = 4), open circles indicate individual observations, closed circle indicates the average and the vertical line indicates the standard error of the mean. **b** ChIP-qPCR analysis of H3K9me2 of wild-type and mutants as indicated. (n = 6, 2 IP experiments performed on the same lysate). Open circles indicate individual observations, closed circle indicates the average and the vertical line indicates the standard error of the mean. Immunoprecipitation performed by Lucía Martín Caballero. **c** RNA PolII-S5P occupancy of four LTRs (representative for the four LTR clusters identified in 3.2.1.3)(n = 4). Open circles indicate individual observations, closed circle indicates the average and the vertical line indicates the standard error of the mean.

Lem2 contributes to transcriptional silencing of heterochromatin (Barrales et al. 2016). Therefore, to investigate if Lem2 also silences transcription of LTRs, I studied the occupancy of actively transcribing RNA polymerase II (RNA PolII-S5P) at LTRs (Figure 3.17c). I observed a subtle increase of PolII at some LTRs; however, this increase may not be statistically significant and probably biologically relevant. Furthermore, I noticed that PolII was not enriched at the heterochromatic *tlh1*$^{+}$ gene, which is transcriptionally regulated by Lem2, as previously shown (Caballero et al. 2021)(Figure 3.17c). The reason for this is not clear but could be due to the low abundance of RNA PolII, at this locus is lowly expressed in general, similar to LTRs. Thus, it appears that the RNA PolII-S5P ChIP results may not be conclusive to determine if Lem2 regulates LTRs transcriptionally or post-transcriptionally.

In summary, I have tested several hypotheses by which Lem2 may regulate the silencing of LTRs. I excluded the possibility that Lem2 affects LTR expression through transcriptional regulation of a known LTR regulator. Consistent with the role of Lem2 in heterochromatin silencing, I found that Lem2 is not enriched at LTRs, largely excluding the possibility that it regulates LTRs through tethering them to the nuclear periphery. Likewise, heterochromatin does not seem to play a direct role in the Lem2-dependent regulation of LTRs.

3.2.4 An unbiased genome-wide approach to identify new LTR regulators and potential Lem2 downstream factors

The hypothesis-driven approach and LTR expression profiling analysis did not reveal factors that function directly up- or downstream of Lem2 with regard to LTR silencing. Therefore, I decided to use an unbiased approach and employed a genome-wide genomic screen to identify factors that regulate the expression of LTRs. Genomic screens have been very successfully applied to identify genes involved in the silencing of heterochro-

matin (Thon & Klar 1992; Ekwall & Ruusala 1994; Thon et al. 1994; Bayne et al. 2014; Barrales et al. 2016).

To perform a screen for mutants with defects in LTR silencing, I needed a readout that can be easily measured and quantified. This can be done with the help of reporter genes, which monitor the silenced state of genomic elements. A commonly used reporter is the *ura4*$^{+}$ gene. Its expression leads to a growth defect when the strain is grown on media containing 5-Fluoroorotic acid (5-FOA), as the *ura4*$^{+}$ gene product converts the 5-FOA into 5-fluorouracil, a toxic compound (Boeke et al. 1984). Here, the reporter gene will be inserted in the vicinity of an LTR, based on the hypothesis that the silenced state of the LTR chromatin environment will be adopted by the reporter gene locus. For the genetic screen, this reporter strain will be crossed with a genome-wide library of deletion mutants of non-essential genes (Kim et al. 2010), using the powerful SGA (synthetic gene array) method that allows conducting genetic crosses in high-throughput using a half-automated setup. Perturbation of LTR silencing will lead to expression of the *ura4*$^{+}$ reporter gene, which in turn results in 5-FOA-based toxicity and as a consequence, a growth defect. Thus, by monitoring the growth of each individual crosses of the reporter strain and the library mutants, this allows me to identify genes involved in LTR silencing.

3.2.4.1 Establishment of a LTR reporter stain

To identify mutants involved in LTR silencing, a reporter strain was initially made by inserting the *ura4*$^{+}$ reporter gene upstream of LTR123. This strain was established by a former Master's student in our laboratory, Sabine Stöcker (Figure 3.18a (hereafter referred to 1xLTR reporter)). Using the growth-based FOA silencing assay, I first examined if the vicinity of LTR123 affects *ura4*$^{+}$ expression in this reporter strain. Serial dilutions of the reporter strain were grown on media with an optimized concentration of 5-FOA

and without 5-FOA; however, the growth is reduced in presence of FOA was comparable with a strain expressing endogenous expression levels of *ura4*$^{+}$ (Figure 3.18b). To evaluate the expression of the reporter genes more quantitatively, I performed RT-qPCR. In agreement with the results by the silencing assay, the expression of the *ura4*$^{+}$ reporter gene was not decreased compared to expression of *ura4*$^{+}$ from the endogenous locus (Figure 3.18b). Thus, the silent state of chromatin at LTR123 does not spread to the up-

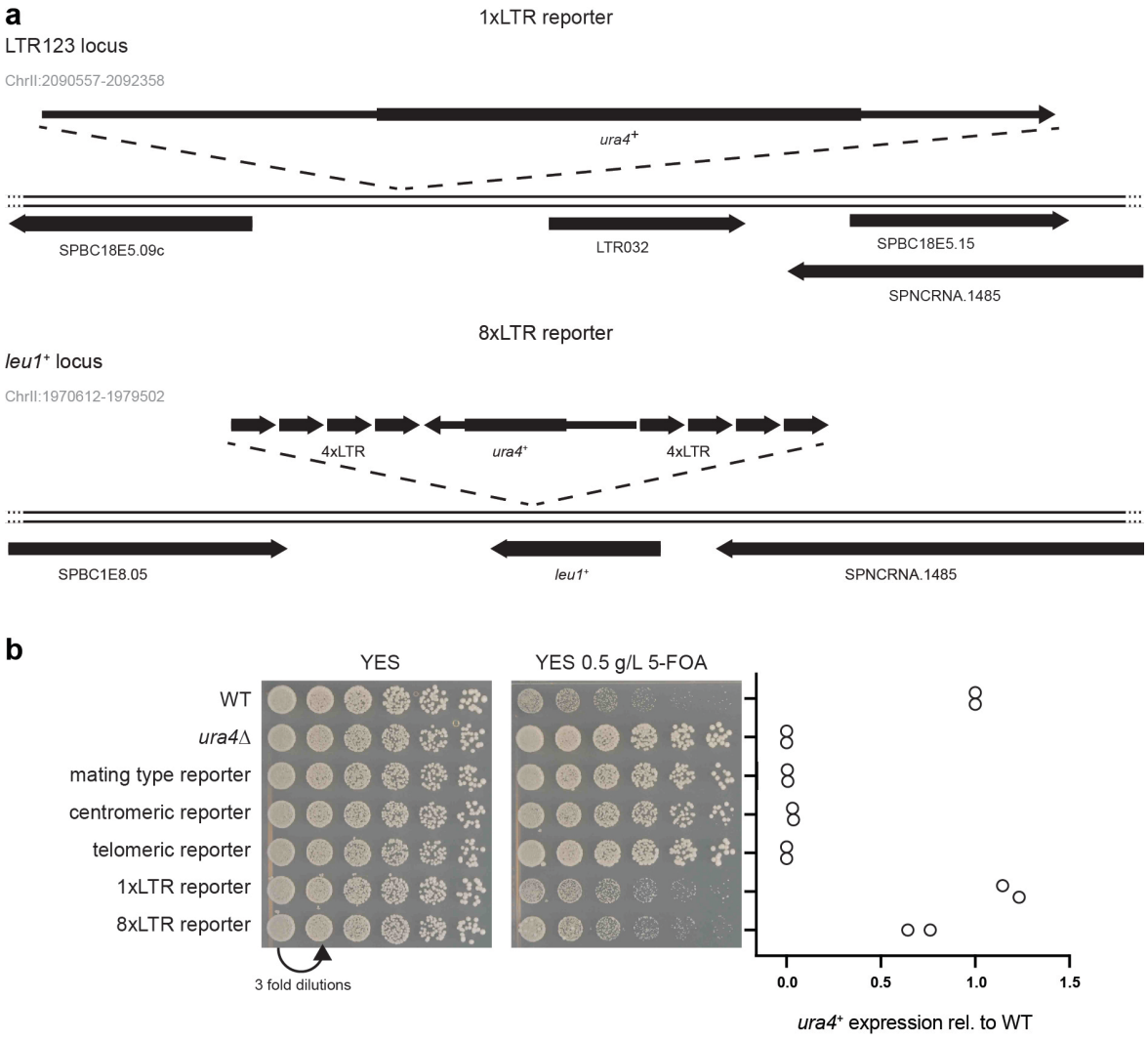


Figure 3.18 Construction of a LTR reporter strains

a Genomic loci where the LTR reporters (1xLTR reporter and 8xLTR reporter) have been inserted. *ura4*$^{+}$ reporter gene and LTRs are drawn to scale. **b** Silencing assay (left) and RT-qPCR (right) to test expression of the *ura4*$^{+}$ reporter gene.

stream inserted *ura4*$^+$ reporter gene.

The lack of 'spreading' of the silent state could caused by the reporter is being not fully embedded into repressed domain. To overcome this limitation the *ura4*$^+$ reporter gene was inserted in a cluster of 8 LTRs (Figure 3.18a (8xLTR reporter)). To increase the chance of identifying mutants that act in the same pathway as Lem2, I selected LTRs that I identified as Lem2-regulated by RNA-seq (Figure 3.15a). Generating constructs containing repetitive DNA elements can be challenging, as PCR reactions cannot correctly amplify repeats. Therefore an elegant cloning approach was used that circumvents the PCR amplification of repetitive DNA (Robinett et al. 1996). This approach is based on recursive digestion of a plasmid with two combinations of restriction enzymes with compatible overhangs but incompatible recognition sequences. After ligation the target sequence is doubled, while one of the restriction sequences in eliminated. This process can be repeated to further expand the repetitive array (see section 2.3.2.2 for more details). The construct after several rounds of digestion and ligation contains a *ura4*$^+$ reporter gene embedded in a cluster of 8 LTRs. This construct was cloned into the *leu1*$^+$ locus, resulting in the 8xLTR reporter strain. To test the expression of the *ura4*$^+$ reporter gene in this reporter strain, I used both growth-based silencing assays and RT-qPCR. The silencing assay showed similar growth on both media types compared to endogenously expressed ura4, similar as observed for the 1xLTR reporter strain. However, the more sensitive RT-qPCR revealed that *ura4*$^+$ expression was decreased by 50% in the 8xLTR reporter strain compared to endogenous *ura4*$^+$. Therefore, the 8xLTR reporter was used to perform a genome-wide genetic screen.

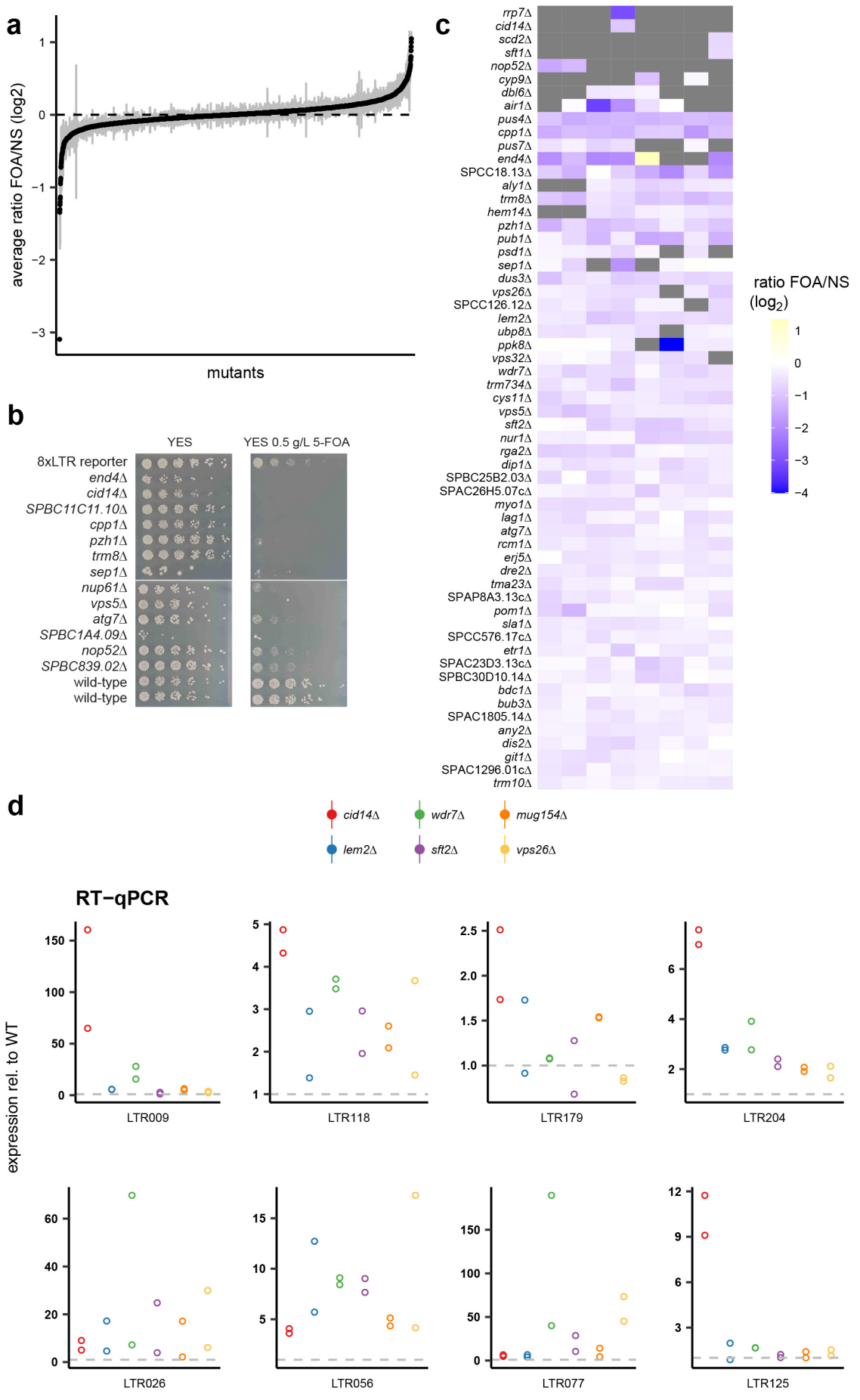
3.2.4.2 A genome-wide genetic screen identifies several LTR regulators

To perform a genetic screen the 8xLTR reporter strain was crossed with a mutant library containing nearly 3000 deletions of non-essential genes (Kim et al. 2010). For each individual genetic cross, the ratio of growth on 5-FOA containing media over growth on non-selective media was calculated and plotted in Figure 3.19a. After repeating this screen four times, each time with two technical replicates, mutants with the lowest average growth ratio are likely candidate LTR regulators. Among those is the mutant of *lem2*$^{+}$, which validates the genetic screening approach.

Therefore, the top 60 candidates, based on ratio of growth on 5-FOA containing media over growth on non-selective media, were tested again in a small-scale growth assay using serial dilutions. For about half of the mutants tested the disrupted silencing of the 8xLTR reporter gene could be confirmed and these mutants were further studied (Figure 3.19b). By examining individual candidates from the genetic screen, it becomes apparent that amongst the top candidates are several mutants with missing values. This is likely caused difficulty of growing mutant strains that are very sick (Figure 3.19c).

It is estimated that 5-10% of the strains in the mutant library are incorrectly annotated. Each stain in the mutant library is marked with a unique barcode, thus this barcode can be used identify the mutants (Kim et al. 2010). Therefore, I sequenced the barcodes of the mutants that were confirmed by the individual serial dilutions experiments. This lead to the exclusion of one mutant from further analysis as the mutation could not be confirmed.

In the genetic screen and the following small scale silencing assay, I measured the growth on 5-FOA containing media as a proxy for LTR expression. To directly assess if the remaining candidates regulate expression of LTRs, I performed RT-qPCR experiments probing the expression of endogenous LTRs, resulting in the conformation of 5 genes


a
average ratio FOA/NS (log2)
mutants
b
YES
YES 0.5 g/L 5-FOA
8xLTR reporter
end4Δ
cid14Δ
SPBC11C11.10Δ
cpp1Δ
pzh1Δ
trm8Δ
sep1Δ
nup61Δ
vps5Δ
atg7Δ
SPBC1A4.09Δ
nop52Δ
SPBC839.02Δ
wild-type
wild-type
c
ratio FOA/NS (log2)
d
cid14Δ
wdr7Δ
mug154Δ
lem2Δ
sft2Δ
vps26Δ
RT−qPCR
expression rel. to WT
LTR009
LTR118
LTR179
LTR204
LTR026
LTR056
LTR077
LTR125

<

Figure 3.19 Genomic screen identifies candidate LTR regulators.

a All mutants of genetic the screen are ranked by average ration of 5-FOA / non-selective media (black). Indicated is also the standard error of the mean (gray). **b** Example image of silencing assay of performed with the reporter strain and 15 deletion library crosses. **c** Heatmap showing the log 2 values of each individual screen for the top 60 mutants. **d** Expression of 8 LTRs (representative for the four LTR clusters identified in 3.2.1.3) in mutants of known factors involved in LTR regulation (n = 2). Open circles indicate individual observations, closed circle indicates the average and the vertical line indicates the standard error of the mean.

(Figure 3.19d). These genes code for the TRAMP complex poly(A) polymerase subunit Cid14, the WD repeat protein orthologous to human WDR7, the Golgi transport protein Sft2, the retromere protein Vps26, and Nur1 that forms a complex with Lem2.

One should keep in mind that any phenotype observed in mutant library strains could potentially be caused by a secondary mutation and not by the annotated mutation. To confirm candidate LTR regulators, I constructed re-knockouts in wild-type background and re-examined expression of a small subset of LTR expression by RT-qPCR. It should be

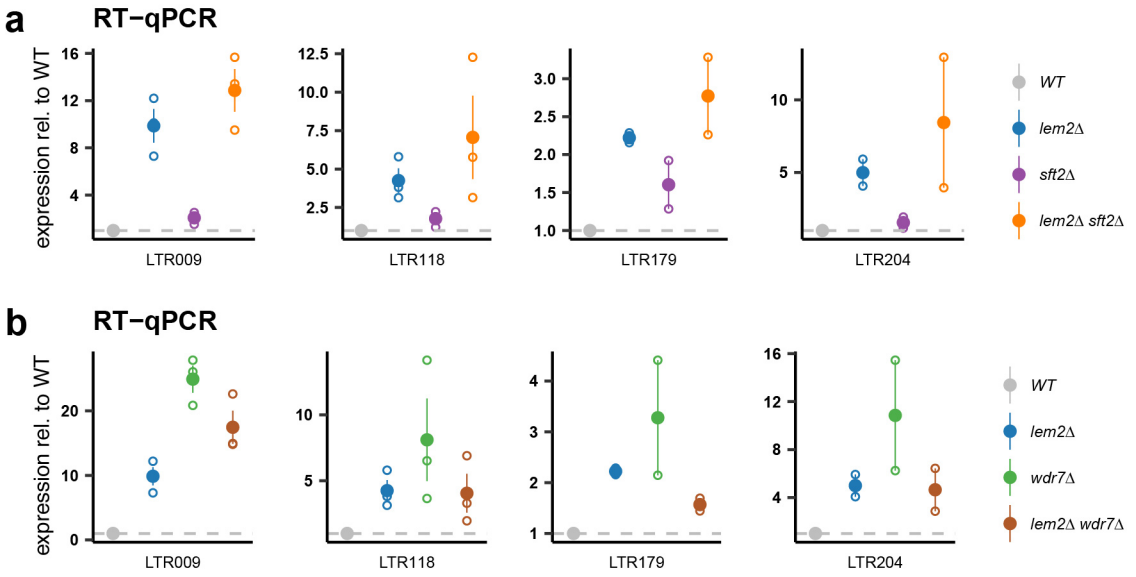


Figure 3.20 *wdr7*$^{+}$ is confirmed as an LTR regulator.

a-b Expression of four LTRs (representative for the four LTR clusters identified in 3.2.1.3) in mutants of known factors involved in LTR regulation (n = 2-3). Open circles indicate individual observations, closed circle indicates the average and the vertical line indicates the standard error of the mean.

noted that this subset does not exclusively contain Lem2 targets. When examining the *sft2*$^+$ re-knockout strain, I detected a minor upregulation in the expression of LTR179 but not for the other LTRs. In contrast, the re-knock-out of *wdr7*$^+$ showed upregulation of LTR expression of all tested LTRs (Figure 3.20ab). This suggests that both *sft2*$^+$ and *wdr7*$^+$ are novel LTR regulators, although the effect of *sft2*$^+$ deletion needs to be further studied as it is only seen in one LTR and the effect is moderate. To examine if Sft2 and Wdr7 work in the same pathway like Lem2, I performed epistasis analysis. The *sft2*Δ *lem2*Δdouble mutant shows a stronger disruption of LTR silencing than the single mutants, suggesting that *sft2*$^+$ acts partially redundantly but does not function in the same pathway as Lem2. With regard to *wdr7*$^+$, although there is large variation in the single mutant, the *wdr7*Δ *lem2*Δ double mutant does not display higher upregulation than the single *lem2*$^+$ mutant, suggesting an epistatic relationship. Further investigation will be needed to elucidate the precise role of these genes. The other screen hits, *cid14*$^+$, *nur1*$^+$ and *vps26*$^+$ have not been subjected to further analysis but represent interesting candidates for follow up studies.

3.2.5 Summary of LTR regulation in *S. pombe*

Studies of full-length retrotransposons have led to the identification of several regulators of these genomic elements. My results show that many of these regulators also regulate solo LTRs. Interestingly, LTR transcription analysis in various mutants shows that LTR regulators have specificity towards subsets of LTRs. However, I also find redundancy between different regulatory pathways.

The inner nuclear membrane protein Lem2 has been known to control subtelomeric LTRs. However, through genome-wide transcription analysis I found that Lem2 is a global regulator of LTRs. The mechanism of Lem2-dependent LTR regulation is unclear.

Through chromatin immunoprecipitation I found that Lem2 does not regulate LTRs through tethering to the nuclear periphery. Instead, through epistasis analysis, I found that the exosome targeting factor Red1 cooperates with Lem2 in the regulation of LTRs, but it remains unclear if these regulators work through the same mechanism. A motif search revealed that the DSR sequence that marks exosome targets is not found in the LTRs I studied, suggesting that these LTRs are targeted through different means.

Finally, by applying a genome-wide genetic screening method I identified several novel LTR regulators, among which are Golgi transport protein Sft2 and WD repeat protein Wdr7. Through epistasis analysis, I found that Sft2 works synthetically with Lem2 and that Wdr7 works in the same pathway as Lem2. How Lem2 coordinates with Red1 and Wdr7 to regulate LTRs provides an interesting scope for further studies.

4. Discussion

4.1 Subtelomeric regulation in *S. pombe*

Even though repetitive DNA makes up a substantial portion of eukaryotic genomes, many aspects of its regulation remain elusive. The subtelomeres of many species contain mosaic repetitive sequence elements (Kwapisz & Morillon 2020). Increasing evidence suggests that these subtelomeric regions play a role in the regulation of telomere maintenance in

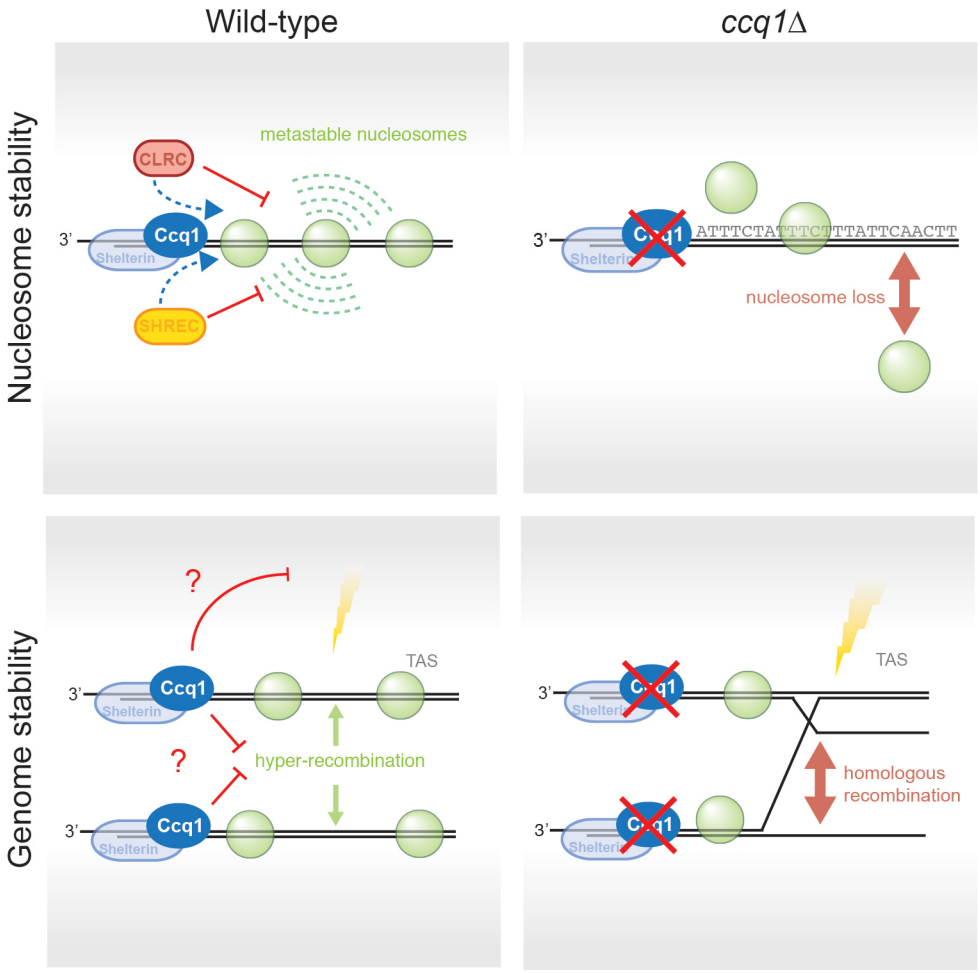


Figure 4.1 Repetitive DNA of subtelomeric TAS regions in *S. pombe* requires Ccq1 for its stability
The DNA sequence of the TAS regions in *S. pombe* are characterized by high A-T content that is refractory to nucleosomes (top right panel). Ccq1 stabilizes nucleosomes the TAS regions through the recruitment of SHREC and CLRC (top left panel, work of Marta Forn). Additionally, Ccq1 inhibits recombination between the homologous TAS regions on different chromosomal arms (bottom panels). This action is however independent of SHREC and CLRC. Figure taken from (van Emden et al. 2019).

the absence of telomerase (Jain et al. 2010). The work described in this dissertation revealed that the TAS regions of *S. pombe* represent a distinct chromatin domain with fundamentally different properties compared to other chromatin domains. The main distinctive features of the *S. pombe* TAS regions are unusually low nucleosome occupancy, likely caused by the high A-T content, and a high recombinogenic potential (Figure 4.1). My findings suggest that these properties of the TAS regions play a role in telomere homeostasis in the absence of telomerase. Here, I discuss potential mechanisms that could underlie the low nucleosome occupancy of the TAS. I also discuss the implications of nucleosome instability for transcription. Finally, I discuss the role of the TAS regions in subtelomeric stability and telomere maintenance.

4.1.1 Possible mechanisms underlying the nucleosome instability at the TAS regions

Results from my co-worker Dr. Marta Forn showed that the H3K9me depletion in the TAS regions of *S. pombe* correlates with, and is likely caused by, the loss of the substrate of lysine 9 methylation, histone H3, as the TAS regions are also characterized by low nucleosome occupancy, which becomes exacerbated in the absence of Ccq1 (van Emden et al. 2019). This provided the foundation of my work, which revealed that TAS fragment inserted at an ectopic site establishes remarkably similar nucleosome levels compared to the endogenous TAS regions (Figure 3.3). This suggests that the low nucleosome density is encoded in the DNA sequence of the TAS regions. Indeed, TAS regions display a high A-T ratio, which has been suggested to be refractory to nucleosomes (Creamer et al. 2014)(Figure 4.1). Thus, the sequence of the TAS regions provides a physical barrier to nucleosome stability and could cause the nucleosome instability that correlates with low H3K9me levels in the TAS region (van Emden et al. 2019).

An alternative explanation for the low nucleosome occupancy of the TAS regions could be competition for DNA binding between nucleosomes and other proteins. One such competitor could be Taz1, component of the shelterin complex. This protein was originally discovered to bind telomeres, but was later also found to bind subtelomeres (Cooper et al. 1997; Kanoh et al. 2005). However, my work shows that deletion of *taz1*$^{+}$ does not restore nucleosome levels in wild-type or *ccq1*Δ cells, but rather causes a further reduction of nucleosome levels in wild-type cells (Figure 3.6).

Another factor that could compete with nucleosomes for TAS binding is the ORC. In fact, ORC requires adenine stretches for binding, which are abundant in the TAS regions (Chuang & Kelly 1999; Okuno et al. 1999). However, adenine stretches themselves are known to be refractory to nucleosomes due to their rigidity (Struhl 1985; Nelson et al. 1987; J. D. Anderson & Widom 2001). Thus it seems more likely that the low nucleosome occupancy allows ORC binding rather than that ORC causes low nucleosome occupancy.

Alternatively, the RSC complex is known to promote nucleosome eviction and has been proposed to counteract the SHREC protein Mit1 (Creamer et al. 2014; Garcia et al. 2010). However, my results suggest that the RSC complex does not play a specific a role in the low nucleosome occupancy of the TAS (Figure 3.5).

Finally, telomeric factors could result in the low nucleosome occupancy observed at the TAS regions, as it has been recently described for telosomes (Greenwood et al. 2018). These distinct, protected structures locate to telomeric DNA and are part of the subtelomere. They are characterized by low nucleosome levels and require telomere binding proteins for their formation (Greenwood et al. 2018). However, the telosome structure does not encroach the subtelomere beyond ~1.5 kb from the telomere, whereas low nucleosome occupancy is observed well beyond this region. Furthermore, the fact that the ectop-

ic TAS fragment has low nucleosome occupancy implies that nucleosome occupancy can be established independent of the telomeric repeats. However, it cannot be excluded that the low nucleosome occupancy at the TAS1 region is partly due to telosome structures, as this region partially overlaps with telosome territory. This could partially explain why the TAS1 region does not have the sharp peaks of high A-T content that are seen in other TAS regions but is nevertheless characterized by low nucleosome occupancy (Figure 3.2 and Figure 3.4).

4.1.2 Transcription of the TAS region

Active transcription requires eviction of nucleosomes to allow RNA polymerase II passing through a gene. Thus, this could also cause low nucleosome occupancy at the TAS regions (Workman 2006). My results show that the reduced nucleosome occupancy of the TAS region in mutants of shelterin, CLRC and SHREC correlate with an increase in TERRA expression (Figure 3.9). SHREC components Mit1 and Clr3 have been shown to play a role in nucleosome stability. Mit1 is a nucleosome remodeler and stabilizes nucleosomes at refractory binding sites, whereas the histone deacetylase Clr3 prevents histone turnover at the silent–mating type region and the pericentromeric regions (Garcia et al. 2010; Aygün et al. 2013; Creamer et al. 2014). However, although silencing is lost at these heterochromatin regions upon deletion of *mit1*$^+$ and *clr3*$^+$, they maintain high levels of nucleosome occupancy in contrast to the TAS regions. This suggests that transcription alone is not sufficient to cause nucleosome instability at the TAS regions.

Interestingly, the silencing of subtelomeric regions beyond the TAS regions, for example at *tlh1*$^+$, also depends on Ccq1 (Figure 3.9). But in contrast to the TAS regions, Ccq1-dependent regulation of *tlh1*$^+$expression is independent of CLRC and SHREC. In agreement with this, the subtelomeric sequences beyond the TAS regions display normal nu-

cleosome abundance. Thus, different regions of the subtelomere are regulated by Ccq1 but through distinct mechanisms. I propose the hypothesis that the appearance of nucleosome refractory sites in the TAS regions coevolved with Ccq1's interaction with CLRC and SHREC in order to maintain a critical level of nucleosomes (van Emden et al. 2019). Nucleosome occupancy at the ectopic TAS region is strikingly similar to the nucleosome occupancy an ectopic TAS regions under the same conditions, both in presence and absence of Ccq1 (Figure 3.3). Thus, an interesting question that has not yet been addressed is if transcription and nucleosome occupancy of ectopic TAS regions are also regulated by CLRC and SHREC recruitment by shelterin. Indeed, discrete blocks of heterochromatin islands that depend on nutritional signals are found across the genome of *S. pombe* (Zofall et al. 2012). Some of these islands require RNA elimination factors, whereas others depend Taz1 and other shelterin components (Zofall et al. 2016). These Taz1-dependent heterochromatin islands rely on heterochromatin for their formation, probably established through CLRC recruitment by shelterin (Zofall et al. 2016). However, in contrast to the TAS region, the heterochromatin island that we tested has marginally decreased nucleosome occupancy in *ccq1*Δ cells (van Emden et al. 2019). This suggests that although shelterin may bind these islands, their regulation happens on the level of chromatin, and not through nucleosome abundance.

4.1.3 The role of the TAS region in subtelomere stability

It has been shown that the TAS region plays an important role in telomere maintenance upon loss of telomerase (Tashiro et al. 2017). An important function of Ccq1 is the recruitment of telomerase (Tomita & Cooper 2008). Most *S. pombe* cells that survive loss of telomerase do so through circularizing their chromosomes. However, upon deletion of *ccq1*$^{+}$, *S. pombe* cells rather survive through genomic rearrangements between the sub-

telomeres (Tomita & Cooper 2008). Such genomic rearrangements of subtelomere have also been observed in this work through the use of subtelomeric reporter genes that function as unique barcodes (Figure 3.10). Upon *ccq1*$^{+}$ deletion the reporter genes become unstable. Importantly, in the absence of the TAS region and other subtelomeric homology regions, these subtelomeric reporter genes are relatively stable (Figure 4.1). While the exact mechanism of recombination remains unknown, it seems likely that break-induced replication plays a role, which has been described in *S. cerevisiae* and humans (McEachern & Haber 2006; Apte & Cooper 2017). In humans this mechanism is called alternative lengthening of telomeres (ALT)(Apte & Cooper 2017). Since the genomic rearrangements are telomerase-independent, this mechanism could be a remnant of an ancient mechanism that evolved prior to telomerase to maintain telomere length.

A remaining interesting question is if the homology between the TAS regions is required for the instability of subtelomeres, or if homology between other regions in subtelomeric arms can provide similar instability. I propose that the properties of the TAS region prime it for homologous recombination, as the high density of poly A-T tracts may be prone to double-strand breaks (Schwartz et al. 2006). Alternatively, specific sequence elements of the TAS could regulate the genomic rearrangements seen in *ccq1*$^{+}$ deletion cells. Such sequence elements may recruit the HR machinery, as it is seen for example in human ALT cells. Here, the nuclear hormone receptor family member NR2C/F recognizes a specific motif and mediates the clustering of repeats to facilitate their recombination (Marzec et al. 2015). Another possibility could be that the TAS region is prone to gross subtelomeric regulation, namely RNA:DNA hybrids, a mechanism also seen in ALT cells. The lncRNA TERRA has been shown to be involved in the formation of these structures, which could contribute to the subtelomeric recombination (Apte & Cooper 2017). However, while deletion of CLRC and SHREC mutants cause an increase in the transcription

of TERRA, this does not lead to increased genomic rearrangements in the subtelomere, making it less likely that TERRA plays a role here.

Overall, the nucleosome instability and high recombinogenic potential are properties that distinguish the TAS regions from other *S. pombe* chromatin regions. I propose that these properties allow the TAS regions play to a role in telomere length control. Further studies in *S. pombe* and other organisms are needed to provide further insights into the functions of mosaic homologous subtelomeric sequences.

4.2 Regulation of LTRs in *S. pombe*

Repetitive DNA also constitutes LTRs. These interspersed repeat are found scattered throughout the genome of *S. pombe*. This work offers an in-depth analysis of how Lem2, an inner nuclear membrane protein, regulates the expression of the LTRs. My results showed that Lem2 regulates these interspersed repeats in a genome-wide manner. I demonstrated that Lem2, for a subset of LTRs, regulates silencing cooperatively with the MTREC subunit, Red1. Furthermore, I showed that heterochromatin does not play a role in the Lem2-mediated silencing of LTRs. Finally, this work identified several novel factors involved in controlling the expression of LTRs. Here, I will discuss LTR regulation by Lem2 and the novel LTR regulators that I identified.

4.2.1 Transcriptional regulation of LTRs in *S. pombe*

Results from my co-worker Dr. Ramón Ramos Barrales, showed that inner nuclear membrane protein Lem2 regulates transcription of subtelomeric solo LTRs (Barrales et al. 2016). Additionally, Lem2 has been implicated in a plethora of other nuclear functions, for example regulating nuclear size and maintaining the boundary between the NE and ER (Kume et al. 2019; Hirano et al. 2020). This raises the question whether deletion of

lem2^{+} results in a loss of silencing that is specific to certain domains or rather that loss of silencing occurs globally. However, although many transcripts are upregulated in *lem2*Δ cells, LTR-neighboring genes are not affected, suggesting that the regulation by Lem2 is very specific (Figure 3.15).

While trying to identify functional pathways using hierarchical cluster analysis between Lem2 and other known LTR regulators, I detected strong correlation of LTR expression between *lem2*Δ and *red1*Δ cells (Figure 3.13). Epistasis analysis revealed that Lem2 and Red1 cooperate in the silencing of a portion of LTRs (Figure 3.13). Recently, a study from our laboratory showed that Lem2 regulates ncRNAs and meiotic transcripts through the exosome targeting complex MTREC (Caballero et al. 2021). In particular, we showed that Lem2 physically interacts with the MTREC component Red1, and that an RNA targeted by the exosome localize to the nuclear periphery in a Lem2-dependent manner (Caballero et al. 2021). Although Lem2 doesn't directly bind to RNA, it promotes the interaction between exosome targeting factors and their targets (Caballero et al. 2021). Some transcripts that are degraded by the nuclear exosome are first recognized through a 6 bp recognition motif known as DSR (determinant of selective removal). Transcripts that are targeted through their DSR sequence often have multiple copies of this motif (Yamashita et al. 2012). This DSR sequence, however, rarely occurs within LTRs, suggesting that LTRs transcripts that are degraded by the nuclear exosome are targeted via a different mechanism (Table 15). Interestingly, *lem2*$^{+}$ and *red1*$^{+}$ cooperate in the regulation of a subset of LTRs, while for other LTRs mutants of these two genes show a suppressive phenotype (Figure 3.13). This may suggest that these LTRs are regulated by a competing factor that can only fully control LTRs when the Lem2-Red1-dependent pathways are completely deleted. In this scenario, one hypothesis is that CENP-B proteins are the competing factor, as they regulate LTRs within TF bodies, whereas the Lem2 and Red1-

dependent pathways most likely regulates LTRs at the nuclear periphery. According to this hypothesis, LTRs are only accessible for CENP-B proteins when Lem2 and Red1 are absent.

CENP-B protein Abp1 and Clr3 have previously been shown to work in the same LTR silencing pathway (Cam et al. 2008). Thus, these two factors were expected to form a functional cluster in the correlation matrix. However, compared to the other values in the matrix, the correlation between *abp1*$^+$ and *clr3*$^+$ is not very strong. A possible explanation for this discrepancy is that I investigated the expression of an unbiased set of LTRs, whereas the study by Cam et al. (2008) focused on full-length retrotransposons. It is possible that, besides the epistatic relation with regard to full-length retrotransposons, Abp1 and Clr3 have a synthetic relation with regard to the regulation of other LTRs, explaining the low correlation observed between *abp1*$^+$ and *clr3*$^+$.

4.2.2 Recognition of LTRs by their regulators

Based on examining a subset of LTRs, they can be grouped into 4 clusters using hierarchical clustering analysis of their regulation by known LTR regulators (Figure 3.14). This raises the question what determines whether an LTR is targeted by a certain regulator or not. *De novo* motif searches in the DNA sequence of LTRs did not reveal characteristic motifs for any of the clusters. Abp1 recognizes two binding motifs in LTRs, and indeed I found that these motifs mostly occur in the Abp1-dependent LTR cluster (Lorenz et al. 2012). However, these binding motifs are not restricted to this cluster (Table 15). Likewise, LTR regulating factors are not restricted to specific clusters, as the correlation matrix reveals overlaps between different clusters. Thus, redundancies between pathways exist. In other words, deletion of a single LTR regulator does not result in the full upregu-

lation of LTRs, as other factors may (partially) take over. This masks the effects of a single deletion, which challenges the search for recognition motifs.

Alternatively, LTRs may be recognized through means other than a recognition sequence within the LTRs. *Wtf* elements often neighbor LTRs and might contain the signal for silencing; however, these elements almost exclusively occur on chromosome III and my results show that LTR regulation is not biased towards particular chromosomes (Figure 3.14). This seems to rule out that LTRs are regulated through recognition of their neighboring *wtf* elements.

4.2.3 Identification of novel LTR regulators

To identify possible downstream factors of Lem2 that regulate LTRs, I employed sensitive reporter assays combined with genetics. I identified several mutants of genes that affect the repression of the reporter gene integrated into a cluster of Lem2-regulated LTRs (Figure 3.19). This suggests that these genes play a direct or indirect role in the regulation of LTR expression.

One of the genes, *nur1*$^{+}$, encodes a protein that forms a complex with Lem2. This complex is reported to fine-tune heterochromatin regulation through SHREC and anti-silencing protein Epe1 (Barrales et al. 2016). However, neither SHREC nor Epe1 is reported to play a major role in the regulation of LTRs through a Lem2-dependent pathway, hence the Lem2-Nur1 complex may regulates LTRs through a different mechanism. In *S. cerevisiae* the Lem2-Nur1 complex homolog is called CLIP (chromatin linkage of inner nuclear membrane proteins) complex. This complex associates with rDNA and tethers it to the nuclear periphery to shield it from recombination among rDNA repeats (Mekhail et al. 2008). However, it seems unlikely that the Lem2-Nur1 complex recruits LTRs to the nuclear periphery through direct binding. While the MSC domain of Lem2 is involved in

regulation of LTRs, this domain does not directly engage in DNA binding but rather in the recruitment of repressor complexes (Barrales et al. 2016). Additionally, I found no evidence for Lem2 binding to LTRs (Figure 3.16). It is however, conceivable that the Lem2-Nur1 complex forms a platform that interacts with a silencing or degradation factor, as it has been shown for SHREC and MTREC (Barrales et al. 2016; Caballero et al. 2021). In order to be silenced, LTRs could be (transiently) recruited to this complex by another factor.

Another interesting factor unveiled by the genetic screen is Cid14, which is part of the TRAMP complex, an exosome targeting complex that contributes to the processing and degradation of snoRNA (Bühler et al. 2008). As discussed above, Lem2 overlaps in the regulation of a subset of LTRs with another exosome targeting factor Red1. However, Cid14 and Red1 are part of independent pathways and it is currently unclear if Lem2 also plays a direct role in the Cid14-dependent pathway. It should be kept in mind that individual hits identified in the genomic screen are not necessarily linked to Lem2.

The third candidate regulator of LTRs is retromer protein Vps26. Interestingly, Vps26 bears homology to the human VPS26 protein that is also part of the retromer complex (Mi et al. 2021). This complex plays a role in the recycling of receptors from endosomes to the trans-Golgi network (Burd & Cullen 2014). It is however not immediately clear how this could relate to the regulation of LTRs.

I showed that the candidate LTR regulator *sft2*$^{+}$ likely does not control LTR silencing in the same pathway as Lem2 (Figure 3.20). Sft2 is predicted to be a Golgi transport protein, and like Vps26, involved in vesicle-mediated transport (Lock et al. 2019). Interestingly, a recent study reports that Sft2 interacts with nuclear membrane proteins Cut11 and Ima1 (Varberg et al. 2020). Cut11 is an integral component of the nuclear membrane, and is often used in microscopy studies to mark the nuclear periphery. Ima1, like Lem2, is an

inner nuclear membrane protein with homology to lamin-associated proteins. However, *S. pombe*, like other unicellular organisms, lacks a nuclear lamina that in other metazoans interacts with repressed chromatin (Amendola & van Steensel 2014). Although interaction with Cut11 and Ima1 is a strong indication that Sft2, like Lem2, plays a role at the inner nuclear membrane, the same study found no interaction between Sft2 and Lem2, furthermore suggesting that Sft2 works in different pathway then Lem2 with regard to LTR silencing (Varberg et al. 2020).

The fifth candidate LTR regulator is the WD repeat protein Wdr7. Interestingly, epistasis analysis suggests that Wdr7 functions in the same pathway as Lem2. Like the other candidates, Wdr7 has not been studied extensively in *S. pombe*, but WD repeat proteins have been topic of investigation in many organisms. Tandem copies of the same or similar motif characterize the amino acid sequence of these proteins. These repeats fold together and form a circularized beta-propeller structure. WD repeat proteins are reported to have many functions, usually they form a binding platform or scaffold that coordinates protein interactions (Smith et al. 1999). One can hypothesize that Wdr7 is a binding platform that facilitates the interaction between Lem2 and its silencing partners SHREC and RNA degradation partner MTREC. An additional factor that also binds to this platform could be responsible for recruiting LTRs.

Overall, I showed that redundancy exists between the different known LTR regulatory pathways. My in-depth investigation of LTR regulation networks revealed that Lem2 works in a functional pathway with exosome targeting factor Red1. Furthermore, I identified WD domain protein Wdr7 as an LTR regulator that functions in the same pathway as Lem2. Further studies are needed to elucidate the exact mechanism by which Lem2 cooperates with Red1 and Wdr7 in order to silence LTRs.

References

Abreu, E. et al., 2010. TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. *Molecular and cellular biology*, 30(12), pp.2971–2982.

Allshire, R.C. & Ekwall, K., 2015. Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. *Cold Spring Harbor perspectives in biology*, 7(7), p.a018770.

Allshire, R.C. & Madhani, H.D., 2018. Ten principles of heterochromatin formation and function. *Nature reviews. Molecular cell biology*, 19(4), pp.229–244.

Alting-Mees, M.A. & Short, J.M., 1989. pBluescript II: gene mapping vectors. *Nucleic acids research*, 17(22), p.9494.

Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403–410.

Amendola, M. & van Steensel, B., 2014. Mechanisms and dynamics of nuclear lamina-genome interactions. *Current opinion in cell biology*, 28, pp.61–68.

Anderson, H.E. et al., 2009. The fission yeast HIRA histone chaperone is required for promoter silencing and the suppression of cryptic antisense transcripts. *Molecular and cellular biology*, 29(18), pp.5158–5167.

Anderson, J.D. & Widom, J., 2001. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Molecular and cellular biology*, 21(11), pp.3830–3839.

Apte, M.S. & Cooper, J.P., 2017. Life and cancer without telomerase: ALT and other strategies for making sure ends (don't) meet. *Critical reviews in biochemistry and molecular biology*, 52(1), pp.57–73.

Aygün, O., Mehta, S. & Grewal, S.I.S., 2013. HDAC-mediated suppression of histone turnover promotes epigenetic stability of heterochromatin. *Nature structural & molecular biology*, 20(5), pp.547–554.

Azzalin, C.M. et al., 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science (New York, N.Y.)*, 318(5851), pp.798–801.

Ågren, J.A. & Wright, S.I., 2011. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 19(6), pp.777–786.

Bah, A. et al., 2012. The telomeric transcriptome of Schizosaccharomyces pombe. *Nucleic acids research*, 40(7), pp.2995–3005.

Bailey, T.L. & Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, 2, pp.28–36.

Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–395.

Bannister, A.J. et al., 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(6824), pp.120–124.

Barrales, R.R. et al., 2016. Control of heterochromatin localization and silencing by the nuclear membrane protein Lem2. *Genes & development*, 30(2), pp.133–148.

Baumann, P. & Cech, T.R., 2000. Protection of telomeres by the Ku protein in fission yeast. A. P. Wolffe, ed. *Molecular biology of the cell*, 11(10), pp.3265–3275.

Bayne, E.H. et al., 2014. A systematic genetic screen identifies new factors influencing centromeric heterochromatin integrity in fission yeast. *Genome biology*, 15(10), pp.481–16.

Bähler, J. et al., 1998. Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast (Chichester, England)*, 14(10), pp.943–951.

Belshaw, R. et al., 2007. Rate of recombinational deletion among human endogenous retroviruses. *Journal of virology*, 81(17), pp.9437–9442.

Boeke, J.D., LaCroute, F. & Fink, G.R., 1984. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular & general genetics : MGG*, 197(2), pp.345–346.

Bowen, N.J. et al., 2003. Retrotransposons and their recognition of pol II promoters: a comprehensive survey of the transposable elements from the complete genome sequence of Schizosaccharomyces pombe. *Genome research*, 13(9), pp.1984–1997.

Brachner, A. & Foisner, R., 2011. Evolvement of LEM proteins as chromatin tethers at the nuclear periphery. *Biochemical Society transactions*, 39(6), pp.1735–1741.

Braun, S. et al., 2011. The Cul4-Ddb1(Cdt)² ubiquitin ligase inhibits invasion of a boundary-associated antisilencing factor into heterochromatin. *Cell*, 144(1), pp.41–54.

Brönner, C. et al., 2017. Accumulation of RNA on chromatin disrupts heterochromatic silencing. *Genome research*, 27(7), pp.1174–1183.

Burd, C. & Cullen, P.J., 2014. Retromer: a master conductor of endosome sorting. *Cold Spring Harbor perspectives in biology*, 6(2), p.a016774.

Bühler, M. & Gasser, S.M., 2009. Silent chromatin at the middle and ends: lessons from yeasts. *The EMBO journal*, 28(15), pp.2149–2161.

Bühler, M. et al., 2007. RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell*, 129(4), pp.707–721.

Bühler, M. et al., 2008. TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. *Nature structural & molecular biology*, 15(10), pp.1015–1023.

Bühler, M., Verdel, A. & Moazed, D., 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell*, 125(5), pp.873–886.

C. elegans Sequencing Consortium, 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science (New York, N.Y.)*, 282(5396), pp.2012–2018.

Caballero, L.M. et al., 2021. The inner nuclear membrane protein Lem2 coordinates RNA degradation at the nuclear periphery. *bioRxiv*, p.2021.05.30.446327. Available at: https://www.biorxiv.org/content/10.1101/2021.05.30.446327v1.

Cam, H.P. et al., 2005. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature genetics*, 37(8), pp.809–819.

Cam, H.P. et al., 2008. Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature*, 451(7177), pp.431–436.

Canela, A. et al., 2007. High-throughput telomere length quantification by FISH and its application to human population studies. *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), pp.5300–5305.

Chaudari, A. & Huberman, J.A., 2012. Identification of two telomere-proximal fission yeast DNA replication origins constrained by nearby cis-acting sequences to replicate in late S phase. *F1000Research*, 1, p.58.

Chen, H.-M., Futcher, B. & Leatherwood, J., 2011. The fission yeast RNA binding protein Mmi1 regulates meiotic genes by controlling intron specific splicing and polyadenylation coupled RNA turnover. *PloS one*, 6(10), p.e26804.

Choudhary, M.N. et al., 2020. Co-opted transposons help perpetuate conserved higher-order chromosomal structures. *Genome biology*, 21(1), pp.16–14.

Chuang, R.Y. & Kelly, T.J., 1999. The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), pp.2656–2661.

Coffin, J.M. et al., 1997. Retroviral Virions and Genomes.

Cooper, J.P. et al., 1997. Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature*, 385(6618), pp.744–747.

Cosby, R.L., Chang, N.-C. & Feschotte, C., 2019. Host-transposon interactions: conflict, cooperation, and cooption. *Genes & development*, 33(17-18), pp.1098–1116.

Creamer, K.M. et al., 2014. The Mi-2 homolog Mit1 actively positions nucleosomes within heterochromatin to suppress transcription. *Molecular and cellular biology*, 34(11), pp.2046–2061.

Dayalu, P. & Albin, R.L., 2015. Huntington disease: pathogenesis and treatment. *Neurologic clinics*, 33(1), pp.101–114.

de Koning, A.P.J. et al., 2011. Repetitive elements may comprise over two-thirds of the human genome. G. P. Copenhaver, ed. *PLoS genetics*, 7(12), p.e1002384.

de Lange, T., 2004. T-loops and the origin of telomeres. *Nature reviews. Molecular cell biology*, 5(4), pp.323–329.

Dehé, P.-M. & Cooper, J.P., 2010. Fission yeast telomeres forecast the end of the crisis. *FEBS letters*, 584(17), pp.3725–3733.

Deng, Z. et al., 2009. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Molecular cell*, 35(4), pp.403–413.

Diehl, A.G., Ouyang, N. & Boyle, A.P., 2020. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nature communications*, 11(1), pp.1796–18.

Dobin, A. et al., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), pp.15–21.

Ehret, C.F. & De Haller, G., 1963. Origin, development, and maturation of organelles and organelle systems of the cell surface in *Paramecium*. *Journal of ultrastructure research*, 23, pp.SUPPL6:1–42.

Ekwall, K. & Ruusala, T., 1994. Mutations in rik1, clr2, clr3 and clr4 genes asymmetrically derepress the silent mating-type loci in fission yeast. *Genetics*, 136(1), pp.53–64.

Ekwall, K., Cranston, G. & Allshire, R.C., 1999. Fission yeast mutants that alleviate transcriptional silencing in centromeric flanking repeats and disrupt chromosome segregation. *Genetics*, 153(3), pp.1153–1169.

Emera, D. & Wagner, G.P., 2012. Transposable element recruitments in the mammalian placenta: impacts

and mechanisms. *Briefings in functional genomics*, 11(4), pp.267–276.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

ENCODE Project Consortium et al., 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), pp.799–816.

Feng, G., Leem, Y.-E. & Levin, H.L., 2013. Transposon integration enhances expression of stress response genes. *Nucleic acids research*, 41(2), pp.775–789.

Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. *Trends in genetics : TIG*, 5(4), pp.103–107.

Fondon, J.W. & Garner, H.R., 2004. Molecular origins of rapid and continuous morphological evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 101(52), pp.18058–18063.

Frank, S., 2014. Treatment of Huntington's disease. *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*, 11(1), pp.153–160.

Garcia, J.F. et al., 2010. Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. *Genes & development*, 24(16), pp.1758–1771.

Garfinkel, D.J. et al., 2005. Ty1 copy number dynamics in Saccharomyces. *Genetics*, 169(4), pp.1845–1857.

Ghidini, M. et al., 2020. Immune-Based Therapies and the Role of Microsatellite Instability in Pancreatic Cancer. *Genes*, 12(1), p.33.

Girard, L. & Freeling, M., 1999. Regulatory changes as a consequence of transposon insertion. *Developmental genetics*, 25(4), pp.291–296.

Grant, C.E., Bailey, T.L. & Noble, W.S., 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*, 27(7), pp.1017–1018.

Graur, D. et al., 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome biology and evolution*, 5(3), pp.578–590.

Greenall, A. et al., 2006. Hip3 interacts with the HIRA proteins Hip1 and Slm9 and is required for transcriptional silencing and accurate chromosome segregation. *The Journal of biological chemistry*, 281(13), pp.8732–8739.

Greenwood, J. & Cooper, J.P., 2012. Non-coding telomeric and subtelomeric transcripts are differentially regulated by telomeric and heterochromatin assembly factors in fission yeast. *Nucleic acids research*, 40(7), pp.2956–2963.

Greenwood, J. et al., 2018. Fission yeast telosomes: non-canonical histone-containing chromatin structures dependent on shelterin and RNA. *Nucleic acids research*, 46(17), pp.8865–8875.

Grewal, S.I., 2000. Transcriptional silencing in fission yeast. *Journal of cellular physiology*, 184(3), pp.311–318.

Haber, J.E. & Louis, E.J., 1998. Minisatellite origins in yeast and humans. *Genomics*, 48(1), pp.132–135.

Haberer, G. et al., 2005. Structure and architecture of the maize genome. *Plant physiology*, 139(4), pp.1612–1624.

Hagan, I. et al., 2016. *Fission Yeast: A Laboratory Manual* I. Hagan et al., eds., New York : Cold Spring

Harbor Laboratory Press.

Hansen, K.R. et al., 2005. Global effects on gene expression in fission yeast by silencing and RNA interference machineries. *Molecular and cellular biology*, 25(2), pp.590–601.

Harland, J.L. et al., 2014. Tpz1-Ccq1 and Tpz1-Poz1 interactions within fission yeast shelterin modulate Ccq1 Thr93 phosphorylation and telomerase recruitment. M. Lichten, ed. *PLoS genetics*, 10(10), p.e1004708.

Heidenreich, B. & Kumar, R., 2017. TERT promoter mutations in telomere biology. *Mutation research*, 771, pp.15–31.

Hentges, P. et al., 2005. Three novel antibiotic marker cassettes for gene disruption and marker switching in Schizosaccharomyces pombe. *Yeast (Chichester, England)*, 22(13), pp.1013–1019.

Hirano, Y. et al., 2020. Lem2 and Lnp1 maintain the membrane boundary between the nuclear envelope and endoplasmic reticulum. *Communications biology*, 3(1), pp.276–14.

Hirano, Y. et al., 2018. Lem2 is retained at the nuclear envelope through its interaction with Bqt4 in fission yeast. *Genes to cells : devoted to molecular & cellular mechanisms*, 23(3), pp.122–135.

Iglesias, N. et al., 2020. Native Chromatin Proteomics Reveals a Role for Specific Nucleoporins in Heterochromatin Organization and Maintenance. *Molecular cell*, 77(1), pp.51–66.e8.

Jain, D. et al., 2010. HAATI survivors replace canonical telomeres with blocks of generic heterochromatin. *Nature*, 467(7312), pp.223–227.

James, T.C. & Elgin, S.C., 1986. Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Molecular and cellular biology*, 6(11), pp.3862–3872.

Jeffreys, A.J., Wilson, V. & Thein, S.L., 1985a. Hypervariable “minisatellite” regions in human DNA. *Nature*, 314(6006), pp.67–73.

Jeffreys, A.J., Wilson, V. & Thein, S.L., 1985b. Individual-specific “fingerprints” of human DNA. *Nature*, 316(6023), pp.76–79.

Job, G. et al., 2016. SHREC Silences Heterochromatin via Distinct Remodeling and Deacetylation Modules. *Molecular cell*, 62(2), pp.207–221.

Kanoh, J. et al., 2005. Telomere binding protein Taz1 establishes Swi6 heterochromatin independently of RNAi at telomeres. *Current biology : CB*, 15(20), pp.1808–1819.

Kapusta, A. et al., 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. H. E. Hoekstra, ed. *PLoS genetics*, 9(4), p.e1003470.

Katoh, I. & Kurata, S.-I., 2013. Association of endogenous retroviruses and long terminal repeats with human disorders. *Frontiers in oncology*, 3, p.234.

Kayser, M., 2017. Forensic use of Y-chromosome DNA: a general overview. *Human genetics*, 136(5), pp.621–635.

Keeney, J.B. & Boeke, J.D., 1994. Efficient targeted integration at leu1-32 and ura4-294 in Schizosaccharomyces pombe. *Genetics*, 136(3), pp.849–856.

Kim, D.-U. et al., 2010. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nature biotechnology*, 28(6), pp.617–623.

Kit, S., 1961. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues.

Journal of molecular biology, 3(6), pp.711–716.

Kizer, K.O. et al., 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and cellular biology*, 25(8), pp.3305–3316.

Kopelman, N.M. et al., 2009. Genomic microsatellites identify shared Jewish ancestry intermediate between Middle Eastern and European populations. *BMC genetics*, 10(1), pp.80–14.

Koressaar, T. & Remm, M., 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics (Oxford, England)*, 23(10), pp.1289–1291.

Kume, K. et al., 2019. Nuclear membrane protein Lem2 regulates nuclear size through membrane flow. *Nature communications*, 10(1), pp.1871–8.

Kunarso, G. et al., 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature genetics*, 42(7), pp.631–634.

Kwapisz, M. & Morillon, A., 2020. Subtelomeric Transcription and its Regulation. *Journal of molecular biology*, 432(15), pp.4199–4219.

Lachner, M. et al., 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410(6824), pp.116–120.

Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), pp.860–921.

Lászik, A. et al., 2000. Automated fluorescent detection of a 10 loci multiplex for paternity testing. *Acta biologica Hungarica*, 51(1), pp.99–105.

Lee, N.N. et al., 2013. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell*, 155(5), pp.1061–1074.

Lee, S.Y. et al., 2020. Dense Transposon Integration Reveals Essential Cleavage and Polyadenylation Factors Promote Heterochromatin Formation. *Cell reports*, 30(8), pp.2686–2698.e8.

Levin, H.L., Weaver, D.C. & Boeke, J.D., 1990. Two related families of retrotransposons from Schizosaccharomyces pombe. *Molecular and cellular biology*, 10(12), pp.6791–6798.

Levinson, G. & Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular biology and evolution*, 4(3), pp.203–221.

Li, B. & Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*, 12(1), pp.323–16.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), pp.2078–2079.

Li, W. et al., 2012. Transposable elements in TDP-43-mediated neurodegenerative disorders. K. M. Iijima, ed. *PloS one*, 7(9), p.e44099.

Lindblad-Toh, K. et al., 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069), pp.803–819.

Lock, A. et al., 2019. PomBase 2018: user-driven reimplementation of the fission yeast database provides rapid and intuitive access to diverse, interconnected information. *Nucleic acids research*, 47(D1), pp.D821–D827.

Lorenz, D.R. et al., 2012. CENP-B cooperates with Set1 in bidirectional transcriptional silencing and ge-

nome organization of retrotransposons. *Molecular and cellular biology*, 32(20), pp.4215–4225.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), pp.550–21.

Luke, B. et al., 2008. The Rat1p 5“ to 3” exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae. *Molecular cell*, 32(4), pp.465–477.

Lynch, H.T. et al., 2009. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clinical genetics*, 76(1), pp.1–18.

Mallet, P.-L., Larochelle, M. & Bachand, F., 2017. Multiple Transcriptional and Post-transcriptional Pathways Collaborate to Control Sense and Antisense RNAs of Tf2 Retroelements in Fission Yeast. *Genetics*, 205(2), pp.621–632.

Martens, J.H.A. et al., 2005. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO journal*, 24(4), pp.800–812.

Martienssen, R. & Moazed, D., 2015. RNAi and heterochromatin assembly. *Cold Spring Harbor perspectives in biology*, 7(8), p.a019323.

Martin, C. & Zhang, Y., 2005. The diverse functions of histone lysine methylation. *Nature reviews. Molecular cell biology*, 6(11), pp.838–849.

Marzec, P. et al., 2015. Nuclear-receptor-mediated telomere insertion leads to genome instability in ALT cancers. *Cell*, 160(5), pp.913–927.

Matsuda, A. et al., 2015. Highly condensed chromatins are formed adjacent to subtelomeric and decondensed silent chromatin in fission yeast. *Nature communications*, 6(1), pp.7753–12.

McEachern, M.J. & Haber, J.E., 2006. Break-induced replication and recombinational telomere elongation in yeast. *Annual review of biochemistry*, 75(1), pp.111–135.

Mejía-Ramírez, E. et al., 2005. The mating type switch-activating protein Sap1 Is required for replication fork arrest at the rRNA genes of fission yeast. *Molecular and cellular biology*, 25(19), pp.8755–8761.

Mekhail, K. et al., 2008. Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature*, 456(7222), pp.667–670.

Mi, H. et al., 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic acids research*, 49(D1), pp.D394–D403.

Miyoshi, T. et al., 2008. Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length. *Science (New York, N.Y.)*, 320(5881), pp.1341–1344.

Moravec, M. et al., 2016. TERRA promotes telomerase-mediated telomere elongation in Schizosaccharomyces pombe. *EMBO reports*, 17(7), pp.999–1012.

Morris, S.A. et al., 2005. Histone H3 K36 methylation is associated with transcription elongation in Schizosaccharomyces pombe. *Eukaryotic cell*, 4(8), pp.1446–1454.

Moser, B.A., Raguimova, O.N. & Nakamura, T.M., 2015. Ccq1-Tpz1TPP1 interaction facilitates telomerase and SHREC association with telomeres in fission yeast. K. S. Bloom, ed. *Molecular biology of the cell*, 26(21), pp.3857–3866.

Mouse Genome Sequencing Consortium et al., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), pp.520–562.

Müller, L.S.M. et al., 2018. Genome organization and DNA accessibility control antigenic variation in

trypanosomes. *Nature*, 563(7729), pp.121–125.

Nakamura, T.M. et al., 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science (New York, N.Y.)*, 277(5328), pp.955–959.

Nakamura, T.M., Cooper, J.P. & Cech, T.R., 1998. Two modes of survival of fission yeast without telomerase. *Science (New York, N.Y.)*, 282(5388), pp.493–496.

Nakayama, J. et al., 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science (New York, N.Y.)*, 292(5514), pp.110–113.

Nelson, H.C. et al., 1987. The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature*, 330(6145), pp.221–226.

Nimmo, E.R. et al., 1998. Defective meiosis in telomere-silencing mutants of Schizosaccharomyces pombe. *Nature*, 392(6678), pp.825–828.

Nimmo, E.R., Cranston, G. & Allshire, R.C., 1994. Telomere-associated chromosome breakage in fission yeast results in variegated expression of adjacent genes. *The EMBO journal*, 13(16), pp.3801–3811.

Noguchi, C. & Noguchi, E., 2007. Sap1 promotes the association of the replication fork protection complex with chromatin and is involved in the replication checkpoint in Schizosaccharomyces pombe. *Genetics*, 175(2), pp.553–566.

Noma K, Allis, C.D. & Grewal, S.I., 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science (New York, N.Y.)*, 293(5532), pp.1150–1155.

Nosek, J., Kosa, P. & Tomaska, L., 2006. On the origin of telomeres: a glimpse at the pre-telomerase world. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 28(2), pp.182–190.

Oizumi, Y. et al., 2021. Complete sequences of Schizosaccharomyces pombe subtelomeres reveal multiple patterns of genome variation. *Nature communications*, 12(1), pp.611–16.

Okuno, Y. et al., 1999. Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. *Molecular and cellular biology*, 19(10), pp.6699–6709.

Orgel, L.E. & Crick, F.H., 1980. Selfish DNA: the ultimate parasite. *Nature*, 284(5757), pp.604–607.

Palm, W. & de Lange, T., 2008. How shelterin protects mammalian telomeres. *Annual review of genetics*, 42(1), pp.301–334.

Panneerchelvam, S. & Norazmi, M.N., 2003. Forensic DNA profiling and database. *The Malaysian journal of medical sciences : MJMS*, 10(2), pp.20–26.

Pearson, C.E., Nichol Edamura, K. & Cleary, J.D., 2005. Repeat instability: mechanisms of dynamic mutations. *Nature reviews. Genetics*, 6(10), pp.729–742.

Persson, J. et al., 2016. Regulating retrotransposon activity through the use of alternative transcription start sites. *EMBO reports*, 17(5), pp.753–768.

Peska, V. & Garcia, S., 2020. Origin, Diversity, and Evolution of Telomere Sequences in Plants. *Frontiers in plant science*, 11, p.117.

Pfeiffer, V. & Lingner, J., 2013. Replication of telomeres and the regulation of telomerase. *Cold Spring Harbor perspectives in biology*, 5(5), pp.a010405–a010405.

Prasanth, S.G. et al., 2004. Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance. *The EMBO journal*, 23(13), pp.2651–2663.

Rea, S. et al., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796), pp.593–599.

Richard, G.-F., Kerrest, A. & Dujon, B., 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and molecular biology reviews : MMBR*, 72(4), pp.686–727.

Robinett, C.C. et al., 1996. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *The Journal of cell biology*, 135(6 Pt 2), pp.1685–1700.

Rodríguez-López, M. et al., 2016. A CRISPR/Cas9-based method and primer design tool for seamless genome editing in fission yeast. *Wellcome open research*, 1(19), p.19.

Rowe, H.M. & Trono, D., 2011. Dynamic control of endogenous retroviruses during development. *Virology*, 411(2), pp.273–287.

Rowe, H.M. et al., 2013. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development (Cambridge, England)*, 140(3), pp.519–529.

Sadaie, M., Naito, T. & Ishikawa, F., 2003. Stable inheritance of telomere chromatin structure and function in the absence of telomeric repeats. *Genes & development*, 17(18), pp.2271–2282.

Schalch, T. & Steiner, F.A., 2017. Structure of centromere chromatin: from nucleosome to chromosomal architecture. *Chromosoma*, 126(4), pp.443–455.

Schmitges, F.W. et al., 2016. Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome research*, 26(12), pp.1742–1752.

Schoeftner, S. & Blasco, M.A., 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nature cell biology*, 10(2), pp.228–236.

Schwartz, M., Zlotorynski, E. & Kerem, B., 2006. The molecular basis of common and rare fragile sites. *Cancer letters*, 232(1), pp.13–26.

Seberg, O. & Petersen, G., 2009. A unified classification system for eukaryotic transposable elements should reflect their phylogeny. *Nature reviews. Genetics*, 10(4), pp.276–276.

Sehgal, A., Lee, C.-Y.S. & Espenshade, P.J., 2007. SREBP controls oxygen-dependent mobilization of retrotransposons in fission yeast. *PLoS genetics*, 3(8), p.e131.

Shahid, S. & Slotkin, R.K., 2020. The current revolution in transposable element biology enabled by long reads. *Current opinion in plant biology*, 54, pp.49–56.

Shimada, K. & Gasser, S.M., 2007. The origin recognition complex functions in sister-chromatid cohesion in Saccharomyces cerevisiae. *Cell*, 128(1), pp.85–99.

Shirasu, K. et al., 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome research*, 10(7), pp.908–915.

Sijen, T. & Plasterk, R.H.A., 2003. Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature*, 426(6964), pp.310–314.

Slotkin, R.K. & Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. *Nature reviews. Genetics*, 8(4), pp.272–285.

Smith, T.F. et al., 1999. The WD repeat: a common architecture for diverse functions. *Trends in biochemical sciences*, 24(5), pp.181–185.

Solovei, I., Gaginskaya, E.R. & Macgregor, H.C., 1994. The arrangement and transcription of telomere

DNA sequences at the ends of lampbrush chromosomes of birds. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 2(6), pp.460–470.

Sotero-Caio, C.G. et al., 2017. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome biology and evolution*, 9(1), pp.161–177.

Sripathy, S.P., Stevens, J. & Schultz, D.C., 2006. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Molecular and cellular biology*, 26(22), pp.8623–8638.

Steglich, B. et al., 2015. The Fun30 chromatin remodeler Fft3 controls nuclear organization and chromatin structure of insulators and subtelomeres in fission yeast. H. D. Madhani, ed. *PLoS genetics*, 11(3), p.e1005101.

Stoye, J.P., 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nature reviews. Microbiology*, 10(6), pp.395–406.

Struhl, K., 1985. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 82(24), pp.8419–8423.

Sugawara, N.F., 1988. *DNA sequences at the telomeres of the fission yeast S. pombe*. thesis, Harvard University, Cambridge, MA.

Sugiyama, T. & Sugioka-Sugiyama, R., 2011. Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast. *The EMBO journal*, 30(6), pp.1027–1039.

Sugiyama, T. et al., 2007. SHREC, an effector complex for heterochromatic transcriptional silencing. *Cell*, 128(3), pp.491–504.

Sundaram, V. et al., 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome research*, 24(12), pp.1963–1976.

Szilard, R.K. et al., 2010. Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX. *Nature structural & molecular biology*, 17(3), pp.299–305.

Takai, K.K. et al., 2011. Telomere protection by TPP1/POT1 requires tethering to TIN2. *Molecular cell*, 44(4), pp.647–659.

Tashiro, S. et al., 2016. Shugoshin forms a specialized chromatin domain at subtelomeres that regulates transcription and replication timing. *Nature communications*, 7(1), pp.10393–12.

Tashiro, S. et al., 2017. Subtelomeres constitute a safeguard for gene expression and chromosome homeostasis. *Nucleic acids research*, 45(18), pp.10333–10349.

The Huntingtons Disease Collaborative Research Group, 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, 72(6), pp.971–983.

Thon, G. & Klar, A.J., 1992. The clr1 locus regulates the expression of the cryptic mating-type loci of fission yeast. *Genetics*, 131(2), pp.287–296.

Thon, G., Cohen, A. & Klar, A.J., 1994. Three additional linkage groups that repress transcription and meiotic recombination in the mating-type region of Schizosaccharomyces pombe. *Genetics*, 138(1), pp.29–38.

Tomita, K. & Cooper, J.P., 2008. Fission yeast Ccq1 is telomerase recruiter and local checkpoint controller. *Genes & development*, 22(24), pp.3461–3474.

Torres-Garcia, S. et al., 2020. SpEDIT: A fast and efficient CRISPR/Cas9 method for fission yeast. *Wellcome open research*, 5(274), p.274.

Vaksman, Z. & Garner, H.R., 2015. Somatic microsatellite variability as a predictive marker for colorectal cancer and liver cancer progression. *Oncotarget*, 6(8), pp.5760–5771.

van Emden, T.S. & Braun, S., 2019. TASks for subtelomeres: when nucleosome loss and genome instability are favored. *Current genetics*, 65(5), pp.1153–1160.

van Emden, T.S. et al., 2019. Shelterin and subtelomeric DNA sequences control nucleosome maintenance and genome stability. *EMBO reports*, 20(1), p.e47181.

Varberg, J.M. et al., 2020. High-Throughput Identification of Nuclear Envelope Protein Interactions in Schizosaccharomyces pombe Using an Arrayed Membrane Yeast-Two Hybrid Library. *G3 (Bethesda, Md.)*, 10(12), pp.4649–4663.

Verstrepen, K.J. et al., 2005. Intragenic tandem repeats generate functional variability. *Nature genetics*, 37(9), pp.986–990.

Vrbsky, J. et al., 2010. siRNA-mediated methylation of Arabidopsis telomeres. T. Kakutani, ed. *PLoS genetics*, 6(6), p.e1000986.

Wang, T. et al., 2007. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proceedings of the National Academy of Sciences of the United States of America*, 104(47), pp.18613–18618.

Weaver, B.A.A. et al., 2007. Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer cell*, 11(1), pp.25–36.

Webb, C.J. & Zakian, V.A., 2012. Schizosaccharomyces pombe Ccq1 and TER1 bind the 14-3-3-like domain of Est1, which promotes and stabilizes telomerase-telomere association. *Genes & development*, 26(1), pp.82–91.

Wells, J.N. & Feschotte, C., 2020. A Field Guide to Eukaryotic Transposable Elements. *Annual review of genetics*, 54, pp.539–561.

Wicker, T. et al., 2007. A unified classification system for eukaryotic transposable elements. *Nature reviews. Genetics*, 8(12), pp.973–982.

Wolf, D. & Goff, S.P., 2007. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell*, 131(1), pp.46–57.

Wolf, D. et al., 2008. Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. *Journal of virology*, 82(9), pp.4675–4679.

Wolf, G. et al., 2015. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes & development*, 29(5), pp.538–554.

Wood, V. et al., 2002. The genome sequence of Schizosaccharomyces pombe. *Nature*, 415(6874), pp.871–880.

Workman, J.L., 2006. Nucleosome displacement in transcription. *Genes & development*, 20(15), pp.2009–2017.

Xi, L. et al., 2010. Predicting nucleosome positioning using a duration Hidden Markov Model. *BMC bioinformatics*, 11(1), p.346.

Xin, H. et al., 2007. TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature*, 445(7127), pp.559–562.

Yamada, K. et al., 2008. Essential roles of Snf21, a Swi2/Snf2 family chromatin remodeler, in fission yeast mitosis. *Genes & genetic systems*, 83(5), pp.361–372.

Yamanaka, S. et al., 2013. RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature*, 493(7433), pp.557–560.

Yamashita, A. et al., 2012. Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. *Open biology*, 2(3), p.120014.

Ye, J. et al., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, 13(1), pp.134–11.

Yun, M. et al., 2011. Readers of histone modifications. *Cell research*, 21(4), pp.564–578.

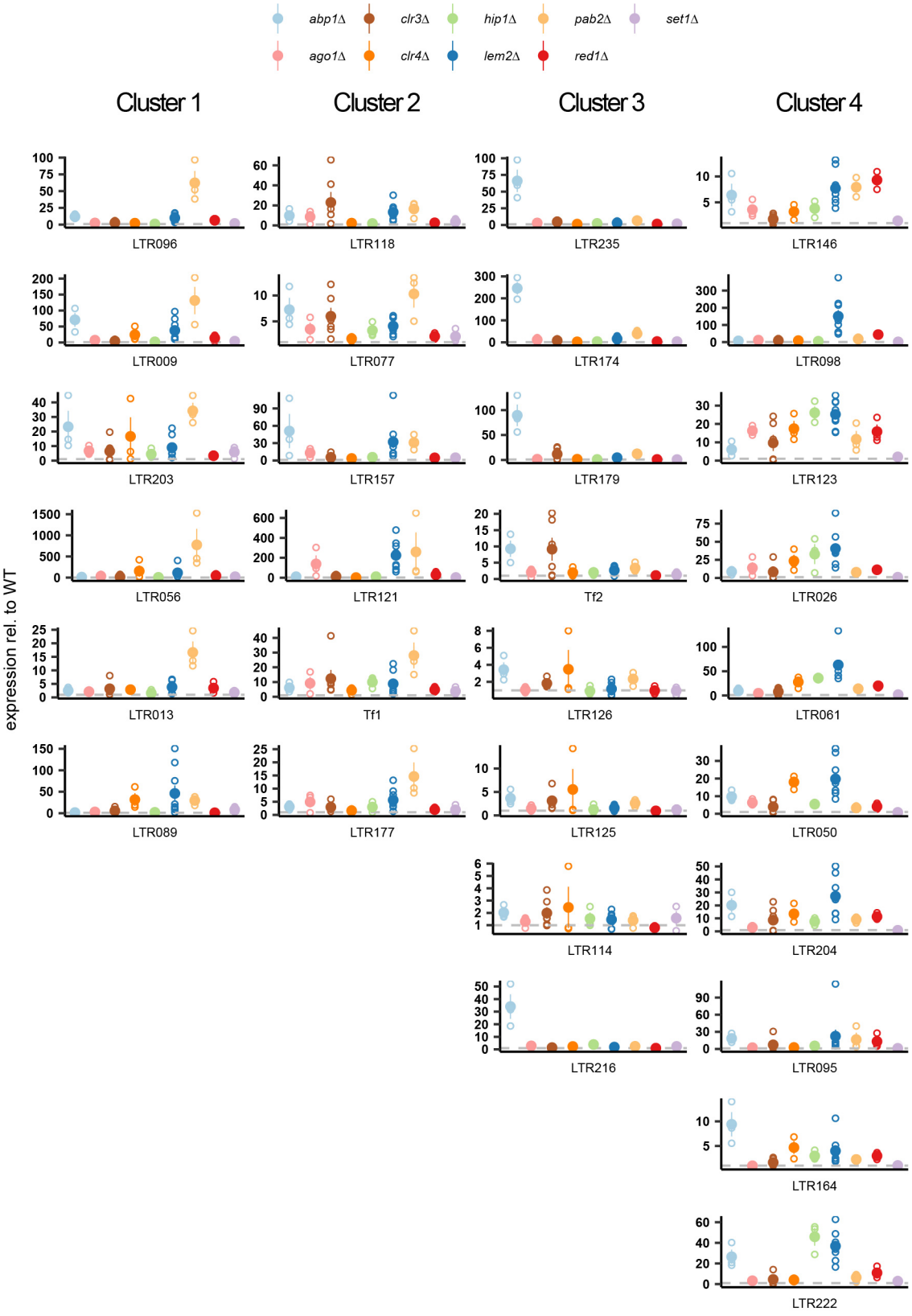
Zaratiegui, M. et al., 2011. CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR. *Nature*, 469(7328), pp.112–115.

Zhang, C.L., McKinsey, T.A. & Olson, E.N., 2002. Association of class II histone deacetylases with heterochromatin protein 1: potential role for histone methylation in control of muscle differentiation. *Molecular and cellular biology*, 22(20), pp.7302–7312.

Zofall, M. et al., 2012. RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science (New York, N.Y.)*, 335(6064), pp.96–100.

Zofall, M. et al., 2016. Taz1-Shelterin Promotes Facultative Heterochromatin Assembly at Chromosome-Internal Sites Containing Late Replication Origins. *Molecular cell*, 62(6), pp.862–874.

Appendix A:


abp1Δ
ago1Δ
clr3Δ
clr4Δ
hip1Δ
lem2Δ
pab2Δ
red1Δ
set1Δ
Cluster 1
Cluster 2
Cluster 3
Cluster 4
expression rel. to WT
LTR096
LTR118
LTR235
LTR146
LTR009
LTR077
LTR174
LTR098
LTR203
LTR157
LTR179
LTR123
LTR056
LTR121
Tf2
LTR026
LTR013
Tf1
LTR126
LTR061
LTR089
LTR177
LTR125
LTR050
LTR114
LTR204
LTR216
LTR095
LTR164
LTR222

<

Appendix A Overview of LTR expression upon which LTRs are grouped

LTR expression in mutants of known factors involved in LTR regulation (n = 9 for *lem2*$^{+}$ and n = 3 for all other mutants). Open circles indicate individual observations, closed circle indicates the average and the vertical line indicates the standard error of the mean. The correlation matrix in Figure 3.14 is based on this data. LTRs are arranged vertically based on the cluster they belong to.

Acknowledgements

First I'd like to thank Andreas Ladurner. As my doktorvater, TAC member and departmental head, he has helped me throughout my PhD journey. Next I'd like to thank Sigurd Braun, my direct supervisor, for giving me the opportunity to perform my doctoral studies in his laboratory. However, as supervisor he did so much more than just having me in the lab. I thank him a lot for helping me grow as a scientist, giving me lots of freedom to take ownership of my projects and being an example as a scientist. I have learned a lot from his critical feedback and have benefitted immensely from his encouragement to dig deeper in my data. I also thank him for giving me the time and freedom to explore the world of programming that I became very passionate about and will continue to explore in the future.

I'd also like to thank the members of the Braun Lab, Sabine, Lucia, Matias, Karl, Abubakar, Agnisrota and Vishnu. They all helped me get through my PhD and have so much fun at the same time. I'm very grateful for their feedback and questions during our lab-meetings as they helped me build my projects and improve my presentations.

During work as a doctoral student I stood on the shoulders of two giants: Marta Forn and Ramon Barrales. I thank Marta for letting me join her project and teaching me the basics of working with *S. pombe,* this solid foundation has helped me throughout my entire PhD. I have really enjoyed continuing Ramon's work on the LTRs and hope to see it published soon. Furthermore, working together with Ramon on the genetic screens has been a great experience.

Lucia and Matias, as members of the Braun Lab, have already been mentioned but in reality, I cannot thank them enough. It has been so great working with them, learning from them and exchange about the challenges we faced. I thank them for all our trips to the

maquina. These little break provided just the right amount of distraction at just the right moments.

I also thank Haris because during my PhD we have bounced ideas back and forth on all kinds of different topics. I really appreciate these interactions, no matter if they were on Angler and Hammerhead or data science careers. Trips to the *maquina* were always extra good with Haris around. Our jointly organized chess club maybe didn't take off as we had hoped but I'm sure we'll get there one day.

I am very grateful to all members of the Ladurner department. My PhD has been very enjoyable and a large part of that is because I was submerged in such a great environment. Ich möchte mich insbesondere bei Marianne und Kathrin bedanken. In den letzten 5 Jahren habe ich ziemlich viele Flaschen YES und noch viel mehr Flasks benutzt. Dank euch waren sie immer da.

I'd also like to thank Lubna and Santiago. It's been a great pleasure supervising these two students. Although I have to admit that they were both very independent so I didn't have to do much supervising. Lubna and Santiago worked on the screen project, the results of which made it into my dissertation. I wish both of them good luck with their academic and medical careers.

For their honest scientific advice I'd like to thank the two TAC members that I have not mentioned yet, Axel Imhof and Guillaume Filion. I am very grateful that they took time out of their busy schedules to advice me on my projects and career goals.

I would also like to thank the bioinformatics facility of the BMC, the CompBio Club and especially Tobias Straub. Through them I have learned a lot about programming and bioinformatics, I hope to continue working on this in my future career.

With all the scientists acknowledged, I would also like to thank the people in my personal environment because I think this work is made possible just as much by them. I thank

Elena, Vanessa Verena and Chris for welcoming me in Munich. Thanks to them I felt at home in Munich from the moment I moved here.

Natuurlijk is niks van al het bovenstaande mogelijk geweest zonder mijn paps en mams. Ik ben dankbaar dat ze elke keuze die ik gemaakt heb richting deze PhD gesteund hebben, ook al betekende het dat ik in het buitenland woon en niet vaak thuis ben. Ik bewonder hun belangstelling voor mijn werk ondanks dat ze het misschien niet helemaal begrijpen.

Finally, I thank Steffi for being there for me, always. I thank her for not complaining when I got up early to go to the lab, or stayed up late to work on this thesis. On the contrary, she has always encouraged me to be great and I very much appreciate that. Together we are embarking on the next big adventure and I look forward to every minute of it.

Affidavit

Affidavit

van Emden, Thomas

Surname, first name

Street

Zip code, town, country

I hereby declare, that the submitted thesis entitled:

Regulation of Transcription and Stability of Repetitive DNA in S. pombe

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

München, 25.05.2022 Thomas van Emden

place, date Signature doctoral candidate