# Subtelomeres as specialized chromatin domains

Antoine Hocher, Angela Taddei

# Subtelomeres as specialized chromatin domains

Antoine Hocher[1,2], Angela Taddei[3,4*]

[1] MRC London Institute of Medical Sciences (LMS), Du Cane Road, London W12 0NN, United Kingdom. [2] Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, United Kingdom

[3] Institut Curie, PSL Research University, CNRS, UMR3664, F-75005 Paris, France. [4] Sorbonne Université, UPMC University Paris 06, CNRS, UMR3664, F-75005 Paris, France

* Corresponding author: angela.taddei@curie.fr

**Keywords:** Subtelomere, Chromosome, Chromatin, Telomere, Histones, Chromatin evolution

**Abstract:**

Specificities associated with chromosomal linearity are not restricted to telomeres. Here, we summarize recent results obtained on fission and budding yeast subtelomeric chromatin and attempt to define subtelomeres using chromatin features extending beyond the heterochromatin emanating from telomeres. Subtelomeres, the chromosome domains adjacent to telomeres, differ from the rest of the genome by their gene content, rapid evolution and chromatin features that together contribute to organism adaptation. However, current definitions of subtelomeres are generally based on synteny and largely gene-centered. Taking into consideration both the peculiar gene content and dynamics as well as the chromatin properties of those domains, we discuss how chromatin features could contribute to subtelomeric properties and functions, and play a pivotal and defining role in the emergence of subtelomeres.
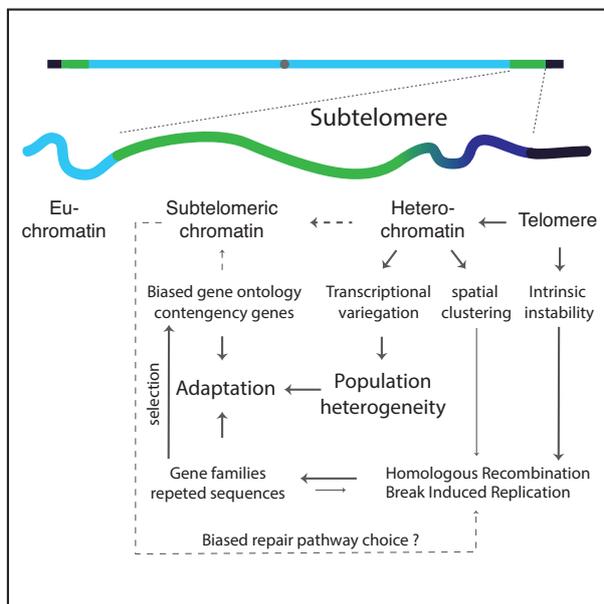
**Graphical abstract:**

Recent work in yeast has highlighted the specificity of subtelomeric chromatin beyond the heterochromatin emerging directly from the telomere. Evidences accumulate that subtelomeric chromatin contributes to the adaptability associated to subtelomeric regions, by impacting both their gene expression and their rapid evolution.

**List of abbreviations:**

- TAS: telomere associated sequences
- ESD: extended Silent Domain
- SIR: Silencing Information Regulators
- TPE: Telomere Position effect

## 1. Introduction

Circular chromosomes reign supreme as they are the favorite genetic storage medium of prokaryotes. Only sporadic examples of linear chromosomes exist in the bacterial kingdom and no linear chromosome has been observed in archaea yet, even though this is likely due to the paucity of work on the subject. In contrast, linear chromosomes, which emerged approximately 1 Gy ago [1], are the prevailing genetic information storage material in eukaryotes. There are a number of reasons that could explain the evolutionary success of linear chromosomes in eukaryotes. Linear chromosomes could be best suited for the maintenance of longer chromosomes [2] and could be advantageous for meiotic division in favoring the spatial clustering of homologous sequences [3]. Alternatively, linear chromosomes could have been a solution to chromosomal instability driven by genome invasion by mobile elements [1]. In this scenario the emergence of telomeres would be a consequence of the invasion of repetitive elements rather than an acquisition providing an instant fitness advantage [4]. Independent of the organism in question or from the potential benefits of the linear topology, linear chromosomes are labor extensive as they come with two direct implications at the cellular level. The first is to differentiate chromosomal ends from a genuine, and accidental, DNA break. The second is to protect linear DNA ends from replication associated erosion, a challenge named "the end replication problem". The structure fulfilling those two functions is commonly referred to as the 'telomere'. Surprisingly, the telomere is not the sole structure associated with chromosomal linearity. While it's difficult to pinpoint their *raison d'être*, subtelomeres, the domains flanking telomeres, often stand out compared to the core genome. No unique properties enable an unambiguous definition of the subtelomere [5], however their functionally and dynamically specific gene content or the occurrence of specific epigenetic processes are characteristic of subtelomeres. This manuscript aims at discussing our current understanding of subtelomeric properties and exploring how the specificity of subtelomeric chromatin could be a defining feature of those domains.

## 2. Delineating telomeres

A necessary prerequisite for the definition of subtelomeres is to first delineate telomeres. Eukaryotic telomeres often consist of short GT-rich tandemly repeated sequences, which are the DNA products of the reverse transcriptase enzyme, telomerase. Because of its nearly ubiquitous presence across eukaryotes, the telomerase is thought to be an ancestral solution to the end replication problem. In many instances however, telomeres cannot merely be reduced to the DNA repeats synthesized by the telomerase. The study of eukaryotes that maintain their

telomeres without telomerase brought valuable insights on telomeric biology. In organisms as diverse as the onion related plants *Allium*, the fruit fly *D. melanogaster* or other insects such as midges, telomeres consist of complex tandem repeat arrays or transposable elements [6]. In *Drosophila*, the most studied of those examples, telomere elongation is also based on RNA reverse transcription, but of two non-LTR (Long terminal Repeats) telomere-specific transposable elements. In this case the end protection and elongation functions of telomeres can be separated. Indeed, cap structures can form in the absence of retrotransposon sequences that are normally found at chromosome ends [7–9]. As a consequence, chromosomes with functional caps may terminate at a variety of locations, and within a variety of sequences [7,10–12]. The fact that telomere capping is independent to the DNA sequence relies on a protein central to heterochromatin formation and maintenance, HP1 (Heterochromatin Protein 1). In *Drosophila*, HP1 directly binds to the telomeric cap and via association with nucleosomes to telomere associated sequences (TAS). Interestingly, telomerase-independent and telomerase-dependent telomeres share several common properties.

The first is the occurrence of epigenetic transcriptional silencing near telomeres, a phenomenon called telomere position effect (TPE). TPE has been thoroughly characterized in budding and fission yeast but is not limited to fungi. Recently two of the three human HP1 isoforms, HP1 alpha and gamma, were shown to have a role in human telomere structure and function [13,14]. In budding yeast, the telomere associated heterochromatin factors such as Sir4, participate in the capping function by inhibiting telomere fusion events [15]. In addition, *S. cerevisiae* heterochromatin factors are also involved in the formation of telomere loops, a well described structure in mammals that remains to be observed by microscopy in budding yeast but can affect gene expression of reporter genes in trans [16,17]. Interestingly, heterochromatin is also central to a class of telomerase mutant survivors in fission yeast, termed HAATI (heterochromatin amplification-mediated and telomerase-independent). The parallel between HAATI and *Drosophila* telomeres suggests a potential ancient role of heterochromatin in telomeric function [18].

The second feature is the presence of DNA repeats that are not synthesized by the telomerase, termed TAS, or of transposable elements near to or embedded within telomeric repeats. Some of those elements, such as the Y' element in budding yeast can be key to survive events of telomerase loss, as amplification by homologous recombination has been observed repeatedly in telomerase mutant survivors [19]. In fission yeast, DNA sequences present in TAS are

intrinsically recombinogenic and promote recombination between chromosomal ends when telomerase recruitment is defective [20].

The involvement of TAS in telomeric functions, albeit infrequent or limited to the rescue of telomerase loss, argues that those elements should be considered as part of the telomere, as suggested previously [21]. In this manuscript, the telomere proximal end of subtelomeres will be operationally defined as the end of TAS repeats when they are present. Importantly, because heterochromatin generally spreads, its nucleation at telomeres already illustrates how certain properties associated to chromosome ends extend further than the telomere, conceptually bringing a direct link between the telomere and the subtelomere.

### 3. Subtelomeres: chromosome domains adjacent to telomeres share specific features

What makes subtelomeres difficult to define is that their function as chromosomal domains remains elusive and the structural properties that set them apart from the core genome can differ considerably among species. In consequence subtelomeres have sometimes been defined by exclusion as non-core and non-telomeric [5]. Here, we will see how current progress on our understanding of subtelomeric chromatin could help in defining subtelomeres. We will first present a non-exhaustive list of properties generally associated to subtelomeres and see how gene content and dynamics have been used to delimit and thus define subtelomeres as discrete domains. Following that, we will elaborate on recent work carried out in both budding and fission yeast that shows how specific chromatin modifiers shape subtelomeres' chromatin landscape and protect euchromatin from the potentially deleterious influence of telomere. Finally, we will discuss the potential mechanisms leading to the emergence of subtelomeric specificities.

*3.1. Topological features*

One direct consequence of chromosomal linearity is the potential to dissipate DNA helical tension by rotation of chromosomal ends. This phenomenon has been elegantly studied in yeast [22]. In this work, DNA was overwound by inactivation of topoisomerases I and II and expression of bacterial TopA. Forced positive supercoiling led to a global down regulation of transcription, except at the chromosomal extremities. Indeed, genes situated at <100 kb from the chromosomal ends gradually escaped from the transcription stall. This effect seems independent of chromatin or telomere structure and largely proportional to the linear distance from the chromosomal end [22]. Although the work on this subject remains limited, there is recent

evidence arguing for a role of the high mobility group protein HMGA2 in regulating supercoiling mediated stability of human subtelomeres [23].

*3.2. Peculiarity of genic content and associated definitions of subtelomeres*

Historically, the first property that has been associated with subtelomeric domains is their particular gene content. Subtelomeres are gene poor and any gene present is generally non-essential. Therefore, essential genes have been proposed as a delimiter for subtelomeres [24], even though this definition is partially context dependent as the essentiality of genes can depend on growth conditions. Another peculiarity of subtelomeres is their enrichment in specific gene families. Pioneering work on the *SUC* and *MAL* gene families in budding yeast stemmed the idea that subtelomeres host gene families that are important for the organism's interaction with its environment [25-27]. Systematic studies later confirmed an enrichment for genes involved in stress response, metabolism of diverse compounds, or transporters [28]. Subtelomeric gene families are larger [28] and show significantly higher copy number variation than their euchromatic counterparts, as revealed by sequencing more than 1000 *S. cerevisiae* isolates [29]. Variable genes being enriched in those regions, subtelomeres are the main contributor to the yeast pan-genome. In addition, budding yeast subtelomeric genes show a higher expression noise (or stochastic fluctuation) than the rest of the genome [30], a property that could be beneficial considering their frequent involvement in stress responses.

Subtelomeric genes are also associated with adaptive processes in other organisms. In various eukaryotic pathogens, including *Trypanosoma brucei*, *Plasmodium falciparum* and *Candida glabrata*, families of adhesin-encoding genes that mediate adhesion to the host cells are localized at subtelomeric regions [31]. The diversity of these genes associated with their epigenetic regulation contribute to the virulence of these microorganisms. Similarly, Olfactory receptor (OR) genes in mammals are clustered and enriched at peri-centromeres and subtelomeres [32]. It has been suggested that the subtelomeric location of OR genes contributes to the diversity of this gene family. In addition to contingency genes, many yeasts and fungi have subtelomere-specific helicases of the RecQ family whose functions are unknown for the most part but their common association with chromosome ends suggests a role for their involvement in telomere maintenance [5]. The peculiarity of subtelomeric gene content can also be exemplified by the fact that many species, including human, host ribosomal DNA arrays within subtelomeres [33]. The different gene content of the subtelomere might not be restricted

to eukaryotes. For example, the subtelomeres of *Streptomyces* bacteria tend to be enriched in gene families involved in secondary metabolism relative to the core genome [2].

Both, the different gene content, and the dynamic character of subtelomeric genes, have independently been used to provide operational definitions of subtelomeres. Comparison of the gene density at the chromosomal end in various Saccharomyces species enabled Brown and colleagues to delimit the subtelomere. Doing so, the average relative genic depletion at chromosomal ends is significant up to 33 kb from the telomere [34]. A drawback of this definition is that the average value does not necessarily account for the variability observed between telomeres and that depends, to some extent, on the set of species used and available for statistical analysis.

### 3.3. Genomic (in)stability

In addition to hosting genes with specialized function, subtelomeres are also a dynamic chromosomal environment. Subtelomeres often differ from the core-genome by the high rate of polymorphism observed between individuals, and by the rapid evolution of these regions between closely related species compared to the core genome [5]. The study of synteny -gene order conservation- among closely related species has been used to delimit subtelomeres. Long-read sequencing of 12 strains representing major subpopulations of *S. cerevisiae* and its wild relative *S. paradoxus* enabled to demarcate the subtelomere from the core genome on the basis of sudden loss of synteny conservation [35]. In the same way, different isolates of the bacteria *Streptomyces* display a zone of progressive synteny degeneration at the border of subtelomeres [36]. In *S. cerevisiae*, this definition efficiently captures distinct properties of chromosomal cores and subtelomeres. Those include the absence of essential genes, a higher accumulation of copy number variants and a higher nonsynonymous-to-synonymous substitution rate ratio (dN/dS) compared to the core-genome. Among the differences between this definition and the one proposed by Brown et al. is the fact that defining the subtelomeres based on synteny offers a delimitation of subtelomere that is specific for each chromosomal arm. Collectively it appears that one can delimit subtelomeres based on the difference of genic structures and dynamics that characterize them. In the following section we will explore the possibility of using epigenetic processes to delimit subtelomeres.

## 3.4. Epigenetic processes at subtelomeres

As mentioned above, the association of heterochromatin factors with telomeres leads to the variegated repression of telomeric proximal genes. The early recognition of the TPE in *Drosophila* demonstrated how genes located in the vicinity of telomeres can experience transcriptional silencing [37]. Extensively studied in budding and fission yeast, TPE leads to the variegated expression of some of the most telomeric proximal subtelomeric genes in several species including mice [38] and humans [39,40]. Given that genes located in subtelomeric regions are often involved in adaptation to the environment, their variegated expression provides a source of cell-to-cell heterogeneity that could play important roles in the capacity of cells to adapt to environmental changes. For example, it was recently shown that the proper regulation of heterochromatin by the conserved Paf1 complex is essential for the long-term survival of fission yeast upon nitrogen starvation [41]. More broadly, heterogeneity may benefit a cell population by allowing escape from commitments, or by implementing a bet hedging strategy for stress resistance. The latter is specifically important for unicellular organisms, and represents a potent mechanism for pathogens to escape the host immune response (e.g. *Plasmodium falciparum*, *Trypanosoma brucei*) or to disseminate in host organisms as shown in *Candida glabrata* [31]. This is exemplified in the malaria agent *P. falciparum* where the ~60 Var genes present in this organism all encode a highly variable transmembrane protein (pfEMP1). However, only a single Var gene is expressed at a time while the remaining ones are epigenetically silenced. By exposing a single variant at a time and switching to a new one when the host has developed an efficient immune response, the parasite maximizes the chances of infection. In the case of *C. glabrata*, the ability to express different kinds of adhesins at specific times during infection determines the ability to persist in the host and colonize different niches [42]. These genes are epigenetically regulated by two nicotinamide adenine dinucleotide (NAD+)-dependent histone deacetylases, Hst1 and Sir2, whose activity can be regulated by the environment. This is well illustrated with the inhibition of Sir2 during urinary tract infections due to low levels of the NAD+ precursor in the environment resulting in the de-repression of the adhesin genes involved in adherence to uroepithelial cells [43]. From a functional point of view, TPE could also be an interesting delimiter of subtelomeres as it emanates from telomeres and spreads towards the chromosomal core. However, regions associated with TPE do not include all the features associated with subtelomeres (see above).

**4. Non-silenced subtelomeric domains in yeast: towards a chromatin-based definition of subtelomeres**

Several elements point toward the idea that subtelomeres are characterized by a particular chromatin, that cannot be reduced to the simple presence of heterochromatin. As mentioned above and detailed in the next section, TPE bridges telomeres to subtelomeres but generally does not occur throughout the whole length of subtelomeres. Other chromosomal features expand above the regions sensitive to TPE, including specific combination of histone marks as shown by independent groups [44,45] and a late replication timing. In the following section, we will detail the current depiction of fission and budding yeast subtelomeres to exemplify how chromatin appears to be an important element of subtelomeric biology (Figure 1).

*4.1. Subtelomeres in yeast*

In *S. pombe*, subtelomeres are composed of distinct parts. Immediately adjacent to most telomeres, the first 10 kb include TAS that are intrinsically refractory to nucleosome formation, show low nucleosome occupancy and low levels of H3K9me2 [20]. Neighboring the TAS, canonical heterochromatin spans over 50 to 80kb while the adjacent telomere-distal subtelomeric chromatin or ST chromatin shows different regional properties than bulk eu- or heterochromatin over ~50 kb [46]. ST chromatin is devoid of the histone H2A variant H2AZ, shows lower levels of histone methylation [44,46] and has higher H3 occupancy than the bulk of the genome. From a cytological point of view, they appear as a highly condensed body or "knob" by super resolution fluorescent microscopy [44]. This condensation is independent of canonical heterochromatin factors such as clr3, dcr1, swi6 or clr4 and does not require H2AZ depletion [44]. In contrast, knob formation requires the H3K36 methyl transferase, Set2, and the centromeric protein, Shugoshin 2 (Sgo2), which is enriched in those domains during the G2 phase [44,47]. The absence of Sgo2 also leads to the de-repression and premature replication of subtelomeric chromatin suggesting that the specific chromatin structure associated with the knob regulates both transcription and replication [47].

The condensation of the knob is dynamic, it disappears in the M phase and reforms in the subsequent G1 phase. Furthermore, the knob does not form upon nitrogen starvation presumably due to the transcriptional activation of ST-chromatin gene in these conditions [44]. Hence, the non-heterochromatic part of *S. pombe* subtelomeres is a dynamic chromosomal domain with specific chromatin features.

In budding yeast, TPE is mediated by the SIR complex. The SIR complex nucleates at TG repeats through interaction with the conserved transcription factor Rap1 that directly binds telomeric repeats. The association of the SIR complex is detected by ChIP-seq at all telomeres in the vicinity to TG repeats, and to the ARS consensus sequence of the X-core elements (ACS), that also act as nucleation sites [48]. Similarly to the TAS element in fission yeast, X-core elements show low nucleosome occupancy [49]. In contrast, the TAS Y' elements show high level of nucleosome occupancy and no SIR enrichment. Consistently, Y' genes are not sensitive to TPE [50] as they are insulated from SIR complex spreading by the binding of the transcription factors Reb1 and Tbf1 to STARS elements [51]. Beyond TAS, Sir3 binding is detected up to 2.6 kb away from the last ACS, on an average [45]. Consistently, TPE affects a rather limited part of subtelomeres in wild-type cells [48].

In contrast, the over representation of gene families or the longer intergenes that are characteristic of subtelomeres extend over larger domains (~30kb on an average) [34]. Complementarily, subtelomeres defined by loss of synteny also extend further away than the wild-type TPE. At last, other properties of chromatin have been shown to differ from the core genome in domains that are larger than the one covered by the SIR complex. For example, histone tail modifications such as H2A phosphorylation (H2AS129ph) are enriched while others, such as H3K79me3, are depleted within domains that extend beyond the domains affected by TPE [45,52]. Discrete subtelomeric subdomains were also revealed upon deletion of chromatin modifiers. This is the case of the HAST domains, which are hyperacetylated in the absence of the histone deacetylase Hda1 [53]. Histone depletion, large deletion of histone H3 tail or point mutation affecting H3 tail acetylation also specifically affect subtelomeric gene expression, well beyond the extent of TPE [54,55].

### 4.2. Expanding heterochromatin reveals discrete subtelomeric domains

Even though heterochromatin silencing only affects a fraction of subtelomeres in wild-type individuals, subtelomeric silent domains can be extended artificially. In budding yeast, increasing the dosage of the silencing factor Sir3 leads to reinforced silencing and extended silent domains covering TAS, including the Y' elements, and subtelomeres up to 19 kb form the last ACS [45,56–58]. Yet, above a certain level, increasing Sir3 levels do not extend silent domains further. Indeed, silent domain extension eventually reaches a plateau upon Sir2 and Sir3 dosage increase. Mapping Sir3 location by chromatin immunoprecipitation demonstrated that Sir3 remains almost entirely restrained to subtelomeres upon overexpression. In this synthetic

configuration, and unlike the wild-type situation, the centromeric proximal ends of silent domains are independent of the silencing nucleation site, *i.e*. the subtelomere length. Those experiments enabled to define the maximal subtelomeric silent domains that were named extended silent domains (ESD). Importantly ESDs not only tell us that heterochromatin spreading has a safety net but also reveal properties of subtelomeric chromatin that are otherwise difficult to detect. ESDs possess a specific chromatin signature in wild-type cells: they are enriched for H2AP, Htz1, and depleted of trimethylated histone H3, whose levels show sharp transitions at the end of ESDs [45]. Accordingly, genes sensitive to the depletion of chromatin modifiers or components such as Htz1, Hda1, Sas2, or the general co-repressor Tup1/Ssn6 are enriched within ESD. Thus, extending heterochromatin to its maximum revealed discrete domains that encompass the features generally associated with subtelomeres. This approach could provide a new way to define subtelomeres that is complementary to the ones based on synteny.

### 4.3. Chromatin and synteny as complementary properties defining subtelomeres
Comparison of ESDs with the location of synteny breakage points brings two slightly different images of what the core genome might look like. At several subtelomeres, it was observed that heterochromatin can spread within the syntenic chromosome core. In addition, some chromatin marks such as H2AS129ph are enriched until the end of ESDs, within what would be defined as a chromosome core by synteny. Accordingly, ESDs segregate histone mark changes along subtelomeres better than synteny-defined subtelomeres' end, which in turn performs better than telomeres defined using a fixed base-pair distance from the telomeres. Satisfyingly, the few (n=3) essential genes that are located within 20kb from telomeres are both classified as euchromatic when considering ESD or synteny-based definitions of subtelomeres. Of course, synteny and chromatin-based delimitation of subtelomeres likely describe different or complementary properties of subtelomeres. One difference probably worth mentioning is that using synteny to delimit subtelomeres is limited to describe what subtelomeres have been 'so-far' as it relies on the comparison of closely related species. Defining subtelomeres using chromatin on the other hand opens the possibility of describing what the subtelomeres are at present. Because of the paucity of the data available on subtelomeric chromatin, this review largely focuses on yeasts. However, we anticipate that the concepts discussed in this review likely apply to other systems.

*4.4. A chromatin tug of war at the end of subtelomeres?*

Euchromatin appears efficiently protected from subtelomeric heterochromatin spreading. As mentioned above, a stable maximal limit of subtelomeric heterochromatin expansion exists in budding yeast [45]. This limit corresponds to zones of chromatin transition that we termed subtelomeric histone mark transition zones. While known punctual barrier elements did not appear to play a major role in constraining silencing, this limit is lost in the absence of the H3K79 methyl-transferase Dot1. Although deleting *DOT1* has little impact on TPE in wild-type cells [59], upon overexpression of Sir3, heterochromatin spreads within euchromatin and leads to synthetic lethality in the absence of Dot1 [45]. Those experiments illustrate the fact that active mechanisms are in place to prevent the aberrant silencing of euchromatic genes (Figure 1). The latter might be particularly important to allow the dynamic variation of subtelomeric size. Indeed, repair events that would position the telomere close to the subtelomeric end can potentially lead to silencing leaking into euchromatin. This situation was studied using fission yeast. It was shown that upon deletion of a large part of subtelomeres, which places telomeres closer to potential synteny break points, heterochromatin over-spreading was prevented by a nucleosome depleted region at one chromosomal arm [60]. Similarly, the maximal extent of subtelomeric heterochromatin was independent of telomere position at telomere 7L in S. cerevisiae [45]. While more work is required to unravel the determinants of the properties of non-heterochromatin subtelomeric domains, they must result from the proximity to the telomere.

*4.5. How does subtelomeric chromatin emerge?*

Chromatin remodelers are required to establish subtelomeric chromatin beyond the domains affected by TPE. In fission yeast, the homolog of chromatin remodeler Fun30, Fft3, localizes at the histone mark transition zone, which in this case corresponds to an array of tandemly repeated LTR transposons [61]. In budding yeast, Ino80c is localized at subtelomeres within domains that partially overlap with SIR bound domains [62,63]. Ino80c localization at subtelomeres is partly dependent on the heterochromatin factor Sir3 [63] and the absence of Ino80 subunit Arp5 leads to elevated transcription in subtelomeres mostly in non-heterochromatic regions. Considering that Sir3 binding is inhibited by H3K79 tri-methylation, the fact that Ino80 inhibits H3K79 methylation by Dot1 *in vitro* [62] could be an important effector of H3K79 hypo-trimethylation at subtelomeres. However, this does not explain why H3 is hypomethylated in subtelomeric regions past the domain covered by Sir3 in wild-type cells. Telomere fusion experiments show that this feature is at least partially independent of the subtelomeric position

of these genes (Hocher and Taddei personal communication). This suggests that the peculiar properties of subtelomeric chromatin could be influenced by the clustering of lowly expressed genes or the nature of gene promoters within these domains.

## 5. How do subtelomeres emerge?

Understanding which phenomenon drives the emergence of subtelomeres certainly is a fascinating question. How does one end up with gene families at the chromosome flanks?

### *5.1. Telomere proximity as the primary source of instability*

Proximity to the end of chromosome *per se* is probably the main driving force as these regions are intrinsically more tolerant than internal sites to rearrangement involving terminal deletion or duplication, simply due to the lower number of genes involved. This alone can lead to the emergence of novel genes by translocation and the emergence of gene families by recombination. The absence of essential genes in these regions further increases their tolerance to mutagenic events, thus favoring the emergence of a high diversity of variants and the accumulation of gene relics. This propensity to duplicate, exchange and alter gene sequences, led to propose that subtelomeres can serve as a nursery for new genes [64]. Furthermore, we recently showed that DSB occurring close to telomeres lead more often to Break Induced Replication (BIR) events due to the rapid loss of a telomeric fragment, which prevents gene conversion [65]. Finally, telomere uncapping events increase the instability of subtelomeres. Dysfunctional telomeres have been proposed to play an adaptive role by increasing the mutation rate of subtelomeres, a process that can result in rapid adaptation to a novel environment [66]. The so-called adaptive telomere failure hypothesis has been recently tested in *K. lactis* using mutation producing stably short telomeres but with no detrimental growth effects on cell growth and viability. This context induces substantially elevated rates of mutation that affect native subtelomeric genes [67]. Surprisingly, half of these mutations corresponds to base changes unaccompanied by gross chromosomal rearrangements. These latter events presumably result from unfaithful Break Induced Replication events between sister chromatids following the uncapping of one of the sisters. These experiments illustrate the potential of mild telomere dysfunction in triggering subtelomere evolution. Interestingly, growing cells for 100 generations at high temperature or under starving conditions lead to telomere shortening in budding yeast possibly accelerating subtelomere evolutions under these stressful conditions [68].

*5.6. Subtelomeric chromatin makes the gene nursery safe and stimulating for creativity*

Processes emanating from telomeres such as heterochromatin or histone tail phosphorylation are also likely to be at the center of what shapes subtelomeres over time. By influencing transcriptional regulation in an environment sensitive manner and within populations, heterochromatin can favor the emergence of adaptive mutations [69]. Secondly, heterochromatin impacts the spatial organization of the genome as it mediates the clustering telomeres and contributes to their perinuclear anchoring [70,71]. In turn, the physical proximity of subtelomeres favor their recombination [65]. In parallel, the association of the Tel1 kinase with telomeres might affect subtelomeric properties. Tel1 presence at the telomere leads to phosphorylation of H2A over large subtelomeric domains. While H2AS129ph is known to contribute to DNA damage signaling and processing [72] and is present at fragile sites [52], its impact on subtelomeres over evolutionary meaningful periods remain to be studied.

In summary, the specific chromatin organization associated with telomere proximity could provide a favorable environment for the gene-nursery associated with subtelomeres, thus contributing to the adaptability of eukaryotes to variable environments (Figure 2). This may be especially relevant for unicellular organisms that have to adapt to variable environments. In particular, eukaryotic pathogens where virulence genes are located in subtelomeres benefit from all the features associated with subtelomeres : (i) their rapid evolution that is essential to generate a large repertoire of genes whose advantageous features can be positively selected for, in response to novel or rare environmental conditions, (ii) the formation of specific chromatin structures allowing the epigenetic regulation of these gene families, which is essential for the colonization of different niches and/or escape from the host immune system. The crucial role of subtelomeric chromatin in the regulation of antigen expression by selection of variant surface glycoprotein genes in *T. brucei*, provides a clear example of such a situation [73].


## 6. Conclusions and outlook

Recent work in yeast has highlighted the specificity of subtelomeric chromatin beyond the heterochromatin emerging directly from the telomere. Evidences accumulate that subtelomeric chromatin contributes to the adaptability associated to subtelomeric regions, by impacting both their gene expression and their rapid evolution. Deciphering the contribution of subtelomeric chromatin to the adaptability of unicellular organisms will have important implications to counteract the virulence of eukaryotic pathogens. Assessing variation in chromatin

organization at the single cell level will help modeling the contribution of epigenetic regulation to the adaptation of a population to a new environment. In parallel, several potentially impactful resources have been made available to study the long-term dynamics of subtelomeres. Experimental evolution [74] of a bacterial strain with artificially engineered linear chromosomes [75] or of a budding yeast strains possessing a single circular or linear chromosome [76,77] could help understanding the causal agents and dynamics of subtelomeric emergence as well as the influence of having subtelomeres on the overall genome dynamics. Similarly, evolution experiments comparing histone point mutants lacking certain histone marks -such as H2AS129ph in budding yeast- could test the contribution of chromatin to the rapid evolution of subtelomeric regions.

Finally, it will be important to expand the current work to more organisms. However, for complex genomes, such as the human genome, single molecule optical mapping technology only recently allowed assembly of subtelomeres and assessment of their variability at the population level [78]. Furthermore, human subtelomeric chromatin remains poorly characterized owing to the difficulties to map short and repetitive DNA elements. Combining long-read sequencing with chromosome conformation data [79] has proved successful to assemble *T. brucei* subtelomeres [73] and should allow rapid progress in our understanding of subtelomeres in the years to come.

**Bibliography:**

[1]    T. De Lange, *Frontiers in Genetics* **2015**, *6*, 321.

[2]    A. Thibessard, P. Leblond, in *Louis E., Becker M. (Eds) Subtelomeres. Springer, Berlin, Heidelberg*, **2014**, pp. 243–258.

[3]    F. Ishikawa, T. Naito, *Mutation Research - DNA Repair* **1999**, *434*, 99.

[4]     N. F. Lue, *Trends in Biochemical Sciences* **2018**, *43*, 314.

[5]     E. Louis, in *Louis E., Becker M. (Eds) Subtelomeres. Springer, Berlin, Heidelberg*, **2014**, pp. 1–12.

[6]     E. J. Louis, *Genome biology* **2002**, *3*, REVIEWS0007.

[7]     R. W. Levis, *Cell* **1989**, *58*, 791.

[8]     R. W. Levis, R. Ganesan, K. Houtchens, L. A. Tolar, F. miin Sheen, *Cell* **1993**, *75*, 1083.

[9]     H. Biessmann, J. M. Mason, K. Ferry, M. d'Hulst, K. Valgeirsdottir, K. L. Traverse, M. Lou Pardue, *Cell* **1990**, *61*, 663.

[10]    K. Ahmad, K. G. Golic, *Genetics* **1998**, *148*, 775.

[11]    S. W. A. Titen, K. G. Golic, *Genetics* **2010**, *184*, 309.

[12]    M. Beaucher, X. F. Zheng, F. Amariei, Y. S. Rong, *Genetics* **2012**, *191*, 407.

[13]    T. T. Chow, X. Shi, J. H. Wei, J. Guan, G. Stadler, B. Huang, E. H. Blackburn, *Nature Communications* **2018**, *9*, DOI 10.1038/s41467-018-05840-y.

[14]    S. Canudas, B. R. Houghtaling, M. Bhanot, G. Sasa, S. A. Savage, A. A. Bertuch, S. Smith, *Genes and Development* **2011**, *25*, 1807.

[15]    S. Marcand, B. Pardo, A. Gratias, S. Cahun, I. Callebaut, *Genes and Development* **2008**, *22*, 1153.

[16]    D. De Bruin, Z. Zaman, R. A. Liberatore, M. Ptashne, *Nature* **2001**, *409*, 109.

[17]    D. de Bruin, S. M. Kantrow, R. A. Liberatore, V. A. Zakian, *Mol Cell Biol* **2000**, *20*, 7991.

[18]    D. Jain, A. K. Hebden, T. M. Nakamura, K. M. Miller, J. P. Cooper, *Nature* **2010**, *467*, 223.

[19]    V. Lundblad, E. H. Blackburn, *Cell* **1993**, *73*, 347.

[20]    T. S. Emden, M. Forn, I. Forné, Z. Sarkadi, M. Capella, L. Martín Caballero, S. Fischer-Burkart, C. Brönner, M. Simonetta, D. Toczyski, M. Halic, A. Imhof, S. Braun, *EMBO reports* **2019**, *20*, DOI 10.15252/embr.201847181.

[21]    J. M. Mason, R. C. Frydrychova, H. Biessmann, *BioEssays* **2008**, *30*, 25.

[22]    R. S. Joshi, B. Pïa, J. Roca, *EMBO Journal* **2010**, *29*, 740.

[23]    S. M. Ahmed, P. D. Ramani, S. Q. Rong Wong, X. Zhao, R. Ivanyi-Nagy, T. C. Leong, C. Chua, Z. Li, H. Hentze, I. B. H. Tan, J. Yan, R. DasGupta, P. Dröge, *PLoS ONE* **2019**, *14*, DOI 10.1371/journal.pone.0215696.

[24]    C. Fairhead, B. Dujon, *FEMS Yeast Research* **2006**, *6*, 428.

[25]    C. A. Michels, E. Read, K. Nat, M. J. Charron, *Yeast* **1992**, *8*, 655.

[26] M. J. Charron, C. A. Michels, *Genetics* **1988**, *120*, 83.

[27] M. J. Charron, E. Read, S. R. Haut, C. A. Michels, *Genetics* **1989**, *122*, 307.

[28] C. A. Brown, A. W. Murray, K. J. Verstrepen, *Current Biology* **2010**, *20*, 895.

[29] J. Peter, M. De Chiara, A. Friedrich, J. X. Yue, D. Pflieger, A. Bergström, A. Sigwalt, B. Barre, K. Freel, A. Llored, C. Cruaud, K. Labadie, J. M. Aury, B. Istace, K. Lebrigand, P. Barbry, S. Engelen, A. Lemainque, P. Wincker, G. Liti, J. Schacherer, *Nature* **2018**, *556*, 339.

[30] N. N. Batada, L. D. Hurst, *Nature Genetics* **2007**, *39*, 945.

[31] A. Juárez-Reyes, I. Castaño, *Current Genetics* **2019**, *65*, 435.

[32] H. C. Mefford, B. J. Trask, *Nature Reviews Genetics* **2002**, *3*, 91.

[33] J. Sochorová, S. Garcia, F. Gálvez, R. Symonová, A. Kovařík, *Chromosoma* **2018**, *127*, 141.

[34] C. A. Brown, A. W. Murray, K. J. Verstrepen, *Current Biology* **2010**, *20*, 895.

[35] J.-X. Yue, J. Li, L. Aigrain, J. Hallin, K. Persson, K. Oliver, A. Bergström, P. Coupland, J. Warringer, M. C. Lagomarsino, G. Fischer, R. Durbin, G. Liti, *Nature Genetics* **2017**, *49*, 913.

[36] F. Choulet, B. Aigle, A. Gallois, S. Mangenot, C. Gerbaud, C. Truong, F. X. Francou, C. Fourrier, M. Guérineau, B. Decaris, V. Barbe, J. L. Pernodet, P. Leblond, *Molecular Biology and Evolution* **2006**, *23*, 2361.

[37] L. L. Wallrath, S. C. R. Elgin, *Genes and Development* **1995**, *9*, 1263.

[38] M. Pedram, C. N. Sprung, Q. Gao, A. W. I. Lo, G. E. Reynolds, J. P. Murnane, *Molecular and Cellular Biology* **2006**, *26*, 1865.

[39] J. A. Baur, Y. Zou, J. W. Shay, W. E. Wright, *Science* **2001**, *292*, 2075.

[40] C. E. Koering, A. Pollice, M. P. Zibella, S. Bauwens, A. Puisieux, M. Brunori, C. Brun, L. Martins, L. Sabatier, J. F. Pulitzer, E. Gilson, *EMBO Reports* **2002**, *3*, 1055.

[41] E. Oya, M. Durand-Dubief, A. Cohen, V. Maksimov, C. Schurra, J. I. Nakayama, R. Weisman, B. Arcangioli, K. Ekwall, *Epigenetics and Chromatin* **2019**, *12*, DOI 10.1186/s13072-019-0292-7.

[42] A. Castano, P. Sanchez, M. T. Llorente, M. Carballo, A. de la Torre, M. J. Munoz, *Sci Total Environ* **2000**, *247*, 337.

[43] R. Domergue, I. Castaño, A. De Las Peñas, M. Zupancic, V. Lockatell, J. R. Hebel, D. Johnson, B. P. Cormack, *Science* **2005**, *308*, 866.

[44] A. Matsuda, Y. Chikashige, D.-Q. Ding, C. Ohtsuki, C. Mori, H. Asakawa, H. Kimura,

T. Haraguchi, Y. Hiraoka, *Nature communications* **2015**, *6*, 7753.

[45] A. Hocher, M. Ruault, P. Kaferle, M. Descrimes, M. Garnier, A. Morillon, A. Taddei, *Genome Research* **2018**, *28*, 1852.

[46] L. Buchanan, M. Durand-Dubief, A. Roguev, C. Sakalar, B. Wilhelm, A. Strålfors, A. Shevchenko, R. Aasland, A. Shevchenko, K. Ekwall, A. F. Stewart, *PLoS Genetics* **2009**, *5*, DOI 10.1371/journal.pgen.1000726.

[47] S. Tashiro, T. Handa, A. Matsuda, T. Ban, T. Takigawa, K. Miyasato, K. Ishii, K. Kugou, K. Ohta, Y. Hiraoka, H. Masukata, J. Kanoh, *Nature Communications* **2016**, *7*, DOI 10.1038/ncomms10393.

[48] A. Ellahi, D. M. Thurtle, J. Rine, *Genetics* **2015**, *200*, 505.

[49] X. Zhu, C. M. Gustafsson, *PLoS ONE* **2009**, *4*, DOI 10.1371/journal.pone.0006363.

[50] F. E. Pryde, E. J. Louis, *EMBO J* **1999**, *18*, 2538.

[51] G. Fourel, E. Revardel, C. E. Koering, E. Gilson, *EMBO J* **1999**, *18*, 2522.

[52] R. K. Szilard, P. E. Jacques, L. Laramee, B. Cheng, S. Galicia, A. R. Bataille, M. Yeung, M. Mendez, M. Bergeron, F. Robert, D. Durocher, *Nat Struct Mol Biol* **2010**, *17*, 299.

[53] D. Robyr, Y. Suka, I. Xenarios, S. K. Kurdistani, A. Wang, N. Suka, M. Grunstein, *Cell* **2002**, *109*, 437.

[54] J. J. Wyrick, F. C. Holstege, E. G. Jennings, H. C. Causton, D. Shore, M. Grunstein, E. S. Lander, R. A. Young, *Nature* **1999**, *402*, 418.

[55] A. M. Martin, D. J. Pouchnik, J. L. Walker, J. J. Wyrick, *Genetics* **2004**, *167*, 1123.

[56] A. Hecht, S. Strahl-Bolsinger, M. Grunstein, *Nature* **1996**, *383*, 92.

[57] H. Renauld, O. M. Aparicio, P. D. Zierath, B. L. Billington, S. K. Chhablani, D. E. Gottschling, *Genes and Development* **1993**, *7*, 1133.

[58] S. Strahl-Bolsinger, A. Hecht, K. Luo, M. Grunstein, *Genes Dev* **1997**, *11*, 83.

[59] Y. H. Takahashi, J. M. Schulze, J. Jackson, T. Hentrich, C. Seidel, S. L. Jaspersen, M. S. Kobor, A. Shilatifard, *Molecular Cell* **2011**, *42*, 118.

[60] S. Tashiro, Y. Nishihara, K. Kugou, K. Ohta, J. Kanoh, *Nucleic Acids Research* **2017**, *45*, 10333.

[61] A. Strålfors, J. Walfridsson, H. Bhuiyan, K. Ekwall, *PLoS Genetics* **2011**, *7*, DOI 10.1371/journal.pgen.1001334.

[62] Y. Xue, C. Van, S. K. Pradhan, T. Su, J. Gehrke, B. G. Kuryan, T. Kitada, A. Vashisht, N. Tran, J. Wohlschlege, C. L. Peterson, S. K. Kurdistani, M. F. Carey, *Genes and Development* **2015**, *29*, 350.

[63] Y. Xue, S. K. Pradhan, F. Sun, C. Chronis, N. Tran, T. Su, C. Van, A. Vashisht, J. Wohlschlegel, C. L. Peterson, H. T. M. Timmers, S. K. Kurdistani, M. F. Carey, *Molecular Cell* **2017**, *67*, 594.

[64] E. V. Linardopoulou, E. M. Williams, C. Fan, Y. Friedman, J.M. Young, B. J. Trask, Nature 2005, 437, 94.

[65] A. Batté, C. Brocas, H. Bordelet, A. Hocher, M. Ruault, A. Adjiri, A. Taddei, K. Dubrana, *The EMBO Journal* **2017**, *36*, 2609.

[66] J. Nosek, *Origin and Evolution of Telomeres*, CRC Press, **2008**.

[67] J. M. O. Mason, M. J. McEachern, *Genetics* **2018**, *208*, 537.

[68] Y. H. Martin Kupiec, *Fungal Genomics & Biology* **2013**, *03*, DOI 10.4172/2165-8056.1000e113.

[69] D. Stajic, L. Perfeito, L. E. T. Jansen, *Nature Ecology and Evolution* **2019**, *3*, 491.

[70] A. Taddei, S. M. Gasser, *Genetics* **2012**, *192*, 107.

[71] H. Funabiki, I. Hagan, S. Uzawa, M. Yanagida, *Journal of Cell Biology* **1993**, *121*, 961.

[72] M. H. Hauer, S. M. Gasser, *Genes and Development* **2017**, DOI 10.1101/gad.307702.117.

[73] L. S. M. Müller, R. O. Cosentino, K. U. Förstner, J. Guizetti, C. Wedel, N. Kaplan, C. J. Janzen, P. Arampatzi, J. Vogel, S. Steinbiss, T. D. Otto, A. E. Saliba, R. P. Sebra, T. N. Siegel, *Nature* **2018**, *563*, 121.

[74] C. Zeyl, *FEMS Yeast Research* **2006**, *6*, 685.

[75] T. Cui, N. Moro-oka, K. Ohsumi, K. Kodama, T. Ohshima, N. Ogasawara, H. Mori, B. Wanner, H. Niki, T. Horiuchi, *EMBO reports* **2007**, *8*, 181.

[76] Y. Shao, N. Lu, Z. Wu, C. Cai, S. Wang, L. L. Zhang, F. Zhou, S. Xiao, L. Liu, X. Zeng, H. Zheng, C. Yang, Z. Zhao, G. Zhao, J. Q. Zhou, X. Xue, Z. Qin, *Nature* **2018**, *560*, 331.

[77] Y. Shao, N. Lu, C. Cai, F. Zhou, S. Wang, Z. Zhao, G. Zhao, J. Q. Zhou, X. Xue, Z. Qin, *Cell Research* **2019**, *29*, 87.

[78] E. Young, H. Z. Abid, P.-Y. Kwok, H. Riethman, M. Xiao, *bioRxiv* **2019**, 728519.

[79] H. Marie-Nelly, M. Marbouty, A. Cournac, J. F. Flot, G. Liti, D. P. Parodi, S. Syan, N. Guillén, A. Margeot, C. Zimmer, R. Koszul, *Nature Communications* **2014**, *5*, 5695.

**Figure 1:** Subtelomere sub domains in fission and budding yeast. This figure depicts the known subtelomeric sub-domains as well as the major contributors to subtelomeric chromatin properties. The antagonism between heterochromatic, subtelomeric and euchromatic chromatin remodelers is schematically depicted
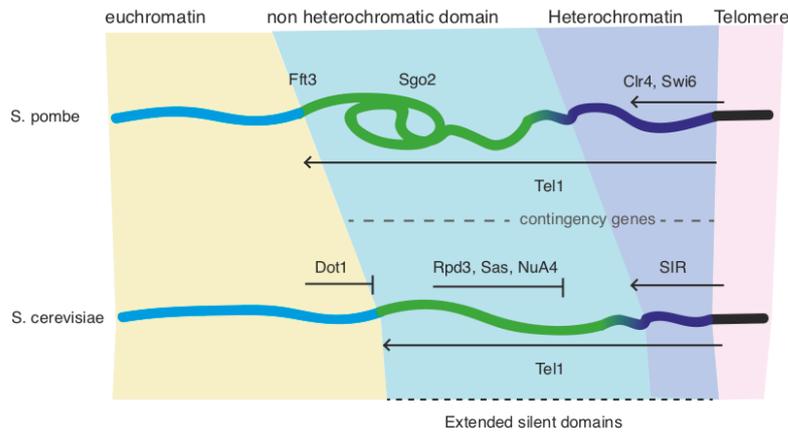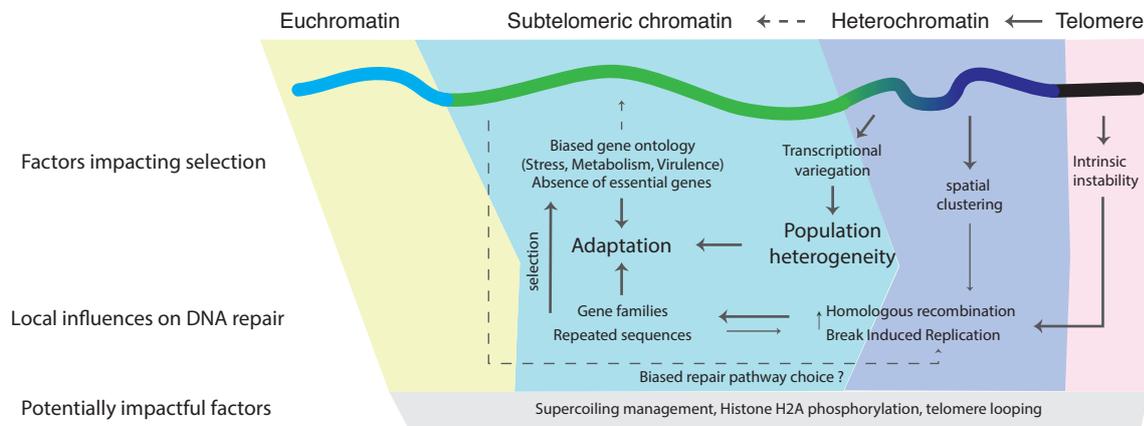


**Fgure 2:** Schematic representation of the interplay between the chromatin and genetic factors contributing to the emergence of subtelomeric properties. Factors whose role have yet to be revealed are mentioned in the bottom grey box.



Hocher and Taddei Figure 2