TÉCNICO LISBOA

TERRA Transcription Profiles in Human Cancer Cells

Miguel Filipe Martins Louro Alves

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisor: Dr. Claus Maria Azzalin

Co-Supervisors: Prof. Cláudia Alexandra Martins Lobato da Silva

Examination Committee

Chairperson: Prof. Maria Margarida Fonseca Rodrigues Diogo

Supervisor: Dr. Claus Maria Azzalin

Member of the Committee: Dr. Simão José Teixeira da Rocha

January 2021

Preface

The work presented in this thesis was performed at the Azzalin lab of the Instituto de Medicina Molecular João Lobo Antunes (Lisbon, Portugal) during the period of January-December 2020, under the supervision of Dr. Claus Maria Azzalin and Dr. Joana Rodrigues. The thesis was co-supervised at Instituto Superior Técnico by Prof. Cláudia Lobato da Silva.

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgements

First of all, I'd like to thank my supervisor, Claus, for all his support, patience and enthusiasm, and for providing me with this great opportunity. I'd like to also thank my co-supervisors, Prof. Cláudia, who was always ready to help with any problem, and Joana, who was incredibly devoted, and helped me overcome every obstacle. A huge thank you to everyone at Azzalin lab for the supportive and kind environment.

Thank you so much to my family and friends for being there for me throughout this tough year of lockdowns and instabilities. And finally, I want to thank my girlfriend Bárbara, who was my rock this year, and was always there to make me laugh, say comforting words, and give me the strength and confidence I needed.

Abstract

TERRA is a long noncoding RNA transcribed from eukaryotic telomeres. It is a major player in the telomerase-independent Alternative Lengthening of Telomeres (ALT) pathway, utilized by approximately 15% of human cancer cells to counteract the telomere shortening that occurs with multiple rounds of cell division, thereby maintaining their immortality. Several ALT cells lines contain elevated levels of TERRA. However, the TERRA transcription profile is still poorly understood, regarding both the levels of TERRA transcription occurring in different cell lines, and the chromosome ends from which TERRA predominantly originates. By long-read sequencing of TERRA, the Azzalin lab found that TERRA levels from chromosome ends 7q and 12q seem to be especially elevated in ALT cells, and hypothesized that quantification of TERRA from these ends could serve as an ALT cancer diagnosis tool. In this project, the TERRA long-read sequencing data was validated by RT-qPCR in a panel of ALT and telomerase-positive cells. TERRA was found to be elevated in U2OS ALT cells at several chromosome ends, particularly 7q, 12q, 16p and 5p. TERRA from these ends was quantified at between 10 and 20 molecules per cell, an average 5-to-10-fold increase compared to telomerase-positive cells. It is proposed that this increase is not sufficient for establishment of a reliable 7q and 12q TERRA diagnosis threshold to discriminate ALT cells. The 7q and 12q ends do not contain the previously characterized CpG dinucleotide-rich TERRA promoter sequences. As such, the second part of this project consisted in the identification, isolation and characterization of a putative new TERRA promoter sequence at these chromosome ends.

Keywords: TERRA, telomere transcription, Alternative Lengthening of Telomeres, RT-qPCR

Resumo

TERRA é um RNA longo não codificante transcrito a partir de telómeros eucariotas. É um desencadeador importante do mecanismo *Alternative Lengthening of Telomeres* (ALT) independente de telomerase, utilizado por aproximadamente 15% dos cancros humanos para contrariar o encurtamento telomérico que ocorre com múltiplas divisões celulares. Várias linhas celulares ALT contêm níveis elevados de TERRA. Contudo, o conhecimento do perfil de transcrição de TERRA continua a ser limitado, tanto em relação aos níveis de transcrição de TERRA em diferentes linhas celulares, como às extremidades cromossómicas a partir das quais TERRA se origina. Através de sequenciação de *reads* de TERRA em células ALT e telomerase-positivas, o laboratório Azzalin verificou que os níveis de TERRA transcrito pelas extremidades cromossómicas 7q e 12q aparentam estar especialmente elevados em células ALT, e levantou a hipótese que a quantificação de TERRA nestas extremidades poderia constituir uma ferramenta diagnóstica de cancros ALT. Neste projeto, os dados da sequenciação de TERRA foram validados através de RT-qPCR. Foi demonstrado que os níveis de TERRA são elevados em várias extremidades, particularmente 7q, 12q, 16p e 5p. TERRA destas extremidades foi quantificado em 10 a 20 moléculas por célula, um aumento médio de 5 a 10 vezes em relação a células telomerase-positivas. Consideramos que este aumento não é suficiente para estabelecer um limiar de TERRA 7q e 12q para discriminar células ALT. As extremidades 7q e 12q não contêm o promotor de TERRA rico em nucleótidos CpG previamente caracterizado. Logo, a segunda parte deste projeto consistiu em identificar, isolar e caracterizar um novo promotor putativo nestas extremidades.

Palavras-Chave: TERRA, transcrição telomérica, *Alternative Lengthening of Telomeres*, RT-qPCR

Table of Contents

List of Figures

List of Tables

List of Acronyms

ALT	Alternative Lengthening of Telomeres
TERRA	telomeric repeat-containing RNA
DDR	DNA damage repair
ssDNA	single-stranded DNA
dsDNA	double-stranded DNA
DSB	double-strand break
TRF1	telomeric repeat binding factors 1
TRF2	telomeric repeat binding factors 2
POT1	protection of telomeres 1
RAP1	repressor and activator protein 1
TIN2	TRF1-interacting nuclear protein 2
TPP1	TIN2 interacting protein
ATM kinase	ataxia telangiectasia mutated kinase
ATR kinase	ataxia telangiectasia and Rad53-related kinase
PARP1	poly(ADP-ribose) polymerase 1
NHEJ	non-homologous end joining
HDR	homology-directed repair
RPA	replication protein A
T-SCEs	telomere sister chromatid exchanges
APBs	ALT-associated PML bodies
PML	promyelocytic leukaemia
ATRX	alpha-thalassemia/mental retardation x-linked
DAXX	death domain-associated protein-6
BIR	break-induced replication
PCNA	proliferating cell nuclear antigen
ATRS	ALT-specific telomeric replication stress
SMARCAL1	helicase SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A-Like 1
FANCM	Fanconi anemia, complementation group M
RNAPII	RNA polymerase II
H4K9	lysine 9 of histone h3
H3K20	lysine 20 of histone h4
NMD	nonsense mediated RNA decay
HP1	heterochromatin protein 1
ONT-seq	Oxford Nanopore Technologies Sequences
RT-qPCR	quantitative reverse transcription polymerase chain reaction
PCR	polymerase chain reaction

cDNA	complementary DNA
gDNA	genomic DNA
CRISPRi	clustered regularly interspaced short palindromic repeats interference
KRAB	Krüppel associated box

1. Introduction

1.1. Motivation

Human chromosomes have ingrained anticancer barriers that suppress unlimited proliferation of cells of the soma, most notably, the telomeric shortening that occurs with multiple rounds of cell replication [1]–[3]. Accordingly, one of the hallmarks of cancer is the sustainment of replicative immortality through counteraction of telomere shortening [4],[5]. In most human cancers, this is achieved by re-expression of a telomere lengthening enzyme known as telomerase. However, a small subset of human cancers, approximately 10-15%, employ a distinct pathway known as Alternative Lengthening of Telomeres (ALT) [6]–[8]. This mechanism elongates telomeres by homology directed repair as a response to telomeric DNA damage. ALT is most common in subsets of sarcomas and astrocytomas [9].

ALT is currently being found in an increasing number of cancer types, including some that have a very poor prognosis with available treatments. Sarcomas that use ALT are less responsive to treatment and are associated to higher risk of patient death compared to telomerase-positive tumors [10]. Therefore, there is a substantial clinical urgency to develop methods for diagnosis and treatment of ALT cancers, an effort that has been hindered by gaps in the current scientific knowledge of this mechanism.

The ALT pathway regulatory circuitry involves an array of ALT activators and alleviators [11]. This thesis focuses on the study of a major player ALT trigger, a long noncoding RNA containing telomeric repeats called TERRA (telomeric repeat-containing RNA). TERRA is transcribed from the subtelomeric region of the chromosome towards the telomere end and anneals with telomeric DNA to form DNA:RNA hybrid structures at telomeres (telR-loops), which play a major role in activating the ALT processes [12].

TERRA has great potential as a target for clinical practice, on both the therapeutic and diagnostic side. TERRA-focused therapies could be critical in treating human ALT cancers. Furthermore, TERRA has been found to be highly upregulated in ALT cells in comparison to telomerase-positive cancer cells [13],[14]. A quantitative study of TERRA expression across different cancer cell lines could potentially open the way to novel ALT cancer diagnosis tools.

TERRA has been shown to be transcribed from several chromosome ends, with some ends potentially expressing considerably more TERRA transcripts than others. Nonetheless, conflicting reports exist on TERRA transcription across different chromosome ends [15]–[18]. The goal of this thesis is to contribute to a better insight into TERRA transcription in human cancer cells, both ALT and non-ALT, which could be used a steppingstone for future development of TERRA-based ALT cancer therapies.

1.2.Telomeres

1.2.1. Structure and Function

Telomeres are heterochromatic nucleoprotein structures that cap the ends of linear eukaryotic chromosomes (Figure 1) and allow cells to distinguish between natural chromosome ends and sites of DNA damage, thus avoiding activation of unwanted DNA damage repair (DDR) processes [19]–[21].

Telomeres are comprised of telomeric DNA, a multitude of telomere-associated proteins (including the shelterin multiprotein complex) and noncoding RNAs. Telomeric DNA is formed by highly conserved 5'-(TTAGGG)-3' tandem repeats in mammals, which extend for 9–15 kb in humans [22],[23].

Most of telomeric DNA is double-stranded (ds). However, at its distal end, the G-rich strand extends beyond its complement to form a G-rich single-stranded (ss) 3'-overhand known as the G-tail [24] (Figure 1). The 3'-overhang is also a general feature of DNA double-strand breaks (DSBs), which means unprotected telomere ends would be recognized as targets for DNA repair machineries, possibly leading to end-to-end telomere fusions. Consequently, telomeric DNA requires a protection mechanism, involving the formation of higher order structures of telomeric DNA known as T-loops (Figure 2B), with the 3' single-stranded overhang looping back and invading the homologous double-stranded region, thereby sequestering the 3' end of chromosomes and preventing its recognition by DNA damage sensors [19],[22],[23].

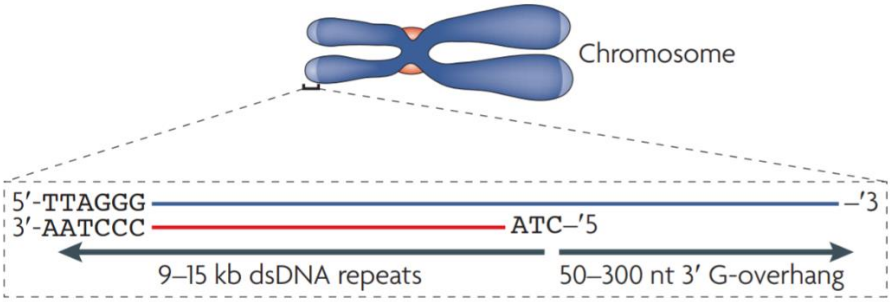

Figure 1: Location of telomeres in the human chromosome and human telomere structure. Human telomeres consist of many kilobases of TTAGGG repeats, with a G-rich leading-strand and a C-rich lagging-strand. The G-strand extends in the 3' direction, forming the G-overhang. Adapted from [20]

Telomeric DNA acts as a docking site for a multiprotein complex called shelterin (Figure 2A). Via its role in in T-loop formation, shelterin is crucial in mediating telomere protection [25],[26]. Shelterin is composed of six distinct individual proteins: telomeric repeat binding factors 1 and 2 (TRF1 and TRF2), repressor and activator protein 1 (RAP1), TRF1-interacting nuclear protein 2 (TIN2), protection of telomeres 1 (POT1), and TIN2 -interacting protein (TPP1)

The shelterin complex assembles at DNA through the action of TRF1 and TRF2, which directly bind to double-stranded telomeric repeats and allow the recruitment of the other shelterin components to chromosome ends [19],[25],[27]. At telomeres, each of the shelterin proteins mediates distinct functions: TRF2 promotes topological changes and T-loop formation and maintenance [28] and represses upstream signaling required for

DNA damage responses, as well as chromosome end fusions resulting from DSB repair pathways [20],[21]. TRF1 has DNA remodeling activity and plays an important role in promoting efficient replication of telomeric DNAs [27]. POT1 and TPP1 associate and bind to the single-stranded 3' overhang where they suppress DNA damage responses and regulate telomerase activity. TIN2 assures the structural integrity of the shelterin complex, as it links the TPP1/POT1 heterodimer to TRF1 and TRF2 and stabilizes TRF1 and TRF2 association to telomeric TTAGGG -repeats [25]. Finally, RAP1 assists in the binding of TRF2 to DNA [29]. The functions of the shelterin complex will be further described in Section 1.2.2.

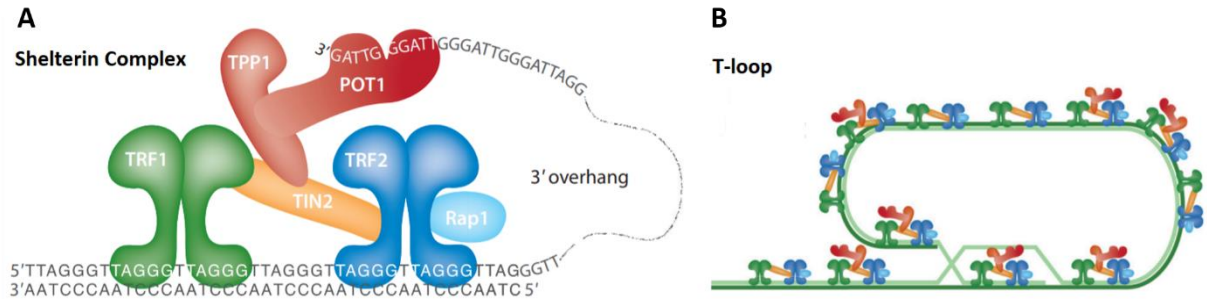

Figure 2: Shelterin structure and function. (A) Depiction of the six-subunit human shelterin complex associated with the double-stranded and single-stranded telomeric DNA. (B) Depiction of shelterin complexes loaded onto telomeres in the t-loop configuration. Adapted from [29]

1.2.2. Chromosome End Protection

As discussed previously, chromosome ends that are unprotected (for example, in absence of a functional shelterin complex) resemble DSBs, and would be recognized as targets for DNA repair [29].

Three distinct DNA damage signaling enzymes, activated by DSBs, are involved in DDR: the ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad53-related (ATR) kinases and poly(ADP-ribose) polymerase 1 (PARP1). ATM or ATR activation by DSBs leads to the phosphorylation of a number of substrates, such as BRCA1, CHK2 and P53, which can lead to downstream processes such as cell cycle arrest and possibly senescence, apoptosis or DNA repair. PARP1 is a potential threat to telomeres because the telomeric DNA sequence contains a PARP1 activation site. PARP1-mediated PARsylation of telomeric proteins, including shelterin subunits, is likely to be detrimental to telomere function because PAR adds negative charges that could interfere with DNA binding and other functions. PARP1 promotes T-loop cleavage and occurrence of DNA repair pathways [29].

Two main DNA repair pathways are involved in the repair of DSBs. First, non-homologous end joining (NHEJ), in which the ends are ligated without the use of homology. NHEJ is divided into classical non-homologous end joining (c-NHEJ) and alternative non-homologous end joining (alt-NHEJ) and it promotes end-to-end fusions of dysfunctional telomeres [30]. Second, homology-directed repair (HDR), which can lead to sequence recombination between two telomeres or a telomere and other parts of the genome [19]. Both pathways are prevented by actions of the shelterin components.

TRF2 represses the downstream activation of the ATM kinase pathway by directing the formation of T-loops, by which the G-overhang is folded into a higher-order complex and concealed. This prevents ATM signaling, which is dependent on the binding of a DSB sensor (the Mre11/Rad50/Nbs11 (MRN) complex) to a free DNA end. TIN2 also contributes to repression of ATM kinase signaling via TRF2 stabilization [29].

As for the ATR kinase pathway, shelterin uses POT1 to repress it [31]. In DSBs across the genome, ATR is stimulated by ssDNA, formed upon resection of the 5' end of a DSB, which is primarily seen during the S phase of the cell cycle. ATR upstream signaling is dependent on the loading of the ssDNA-binding factor replication protein A (RPA) [19],[30]. It is believed that shelterin averts ATR activation at telomeres by RPA exclusion, resultant from competition between POT1 and RPA for ssDNA [31]. Due to its tethering to the shelterin core, POT1 has a binding advantage over RPA to accumulate in excess at the telomeric single-stranded target sites, including the G-overhang, thus blocking RPA access [19],[30]. TPP1 is also needed for repression of the ATM kinase signaling, as POT1 can only perform its inhibitory function when tethered to the rest of the shelterin complex by TPP1 [30], [31]. The binding of PARP1 to telomeres is prevented by TRF2 and TIN2 [29]. It is believed that TRF2 masks the PARP1 activation site using its basic branched-DNA binding domain [31].

TRF2 and POT1 shelterin subunits both contribute to inhibition of the NHEJ and HDR pathways [29]. At DSBs, NHEJ is mediated by the Ku70/80 heterodimer, which is a ring-shaped complex that loads onto accessible DNA ends and brings them together, promoting the ligation of DSBs via DNA ligase IV (Lig4), in complex with other factors[26], [32], [33]. Telomere-telomere fusions through NHEJ leads to formation of dicentric chromosomes. These are a great threat to genome integrity due to their instability during mitosis [21]. TRF2 primarily blocks NHEJ activation by promoting T-loop formation, since their sequestering of the G-overhang prevents loading of the Ku70/80 ring onto it. Depletion of TRF2, or TRF2 stabilizing TIN2, is proven to suppress T-loop formation and provide the substrate needed for Ku70/80 loading and NHEJ stimulation[20],[32]. RAP1 was proposed to be involved in the repression of NHEJ as well, based on a tethering experiment where artificial localization of RAP1 at telomeres devoid of TRF2 limited telomere-telomere fusions [35]. POT1 mediated suppression of NHEJ, in the absence of T-loops, by blocking the efficient loading of Ku70/80, or by preventing cleavage when bound to the G-overhang [36].

As for HDR, how it is repressed at telomeres is still uncertain. HDR between sister telomeres results in telomere sister chromatid exchanges (T-SCEs) [29]. Loss of POT1, TRF2 and RAP1 lead to an increased frequency of T-SCEs in metaphase spreads, suggesting that they suppress HDR [29], [37]–[39]. The mechanism by which they act and the nature of the HDR reaction they are blocking need to be clarified. The suppressing ability of POT1 may rely on its inherent competition for telomeric ssDNA with other ssDNA-binding proteins, which includes not only RPA but also the HDR factor Rad5 [21]. The Ku70/80 heterodimer, due to its role in the NHEJ pathway, also acts to suppress the HDR pathway at DSBs and at telomeres and it does so likely not by loading to the terminal regions but rather by interacting with the shelterin complex [19],[26].

1.2.3. The End-Replication Problem

The semiconservative DNA replication machinery is unable to fully replicate the extremities of linear DNA, due to the failure of lagging-strand synthesis to fully replicate the parental strand. In somatic cells that lack telomere length maintenance mechanisms, this results in progressive telomere attrition, i. e. loss of telomeric sequence, with each round of replication, a phenomenon known as the end replication problem [1], [20],[40].

A brief explanation of the end-replication problem follows. DNA Replication always occurs in the 5' to 3' direction due to presence of 3'-OH groups at nucleotides, these being the site for nucleotide addition, required by DNA polymerases to perform DNA synthesis de novo. The two DNA strands are anti-parallel; there is a leading-strand that runs in the 5' to 3' direction compared to replication fork progression and is replicated continuously. The lagging-strand, however, runs in the 3' to 5' direction towards the replication fork and must be replicated discontinuously. For this process, short RNA primers, made by RNA primase, anneal to the strand [41]. These are then extended by DNA polymerase to form DNA fragments called Okazaki fragments. The primers provide the free 3'-OH group that the replicative DNA polymerase needs to bind and start C-rich strand elongation through the formation and joining of Okazaki fragments, forming the DNA daughter strand [42].

At telomeres, the G-strand serves as a template for lagging-strand synthesis of the C-rich daughter strand, while the C-strand serves as a template for leading-strand synthesis of the G-rich daughter strand. When the RNA primers used in lagging C-strand synthesis are removed, there is no way to synthesize lagging-strand sequence that is complementary to the small region at the end of the chromosome where the terminal primer was degraded. This leads to a loss of nucleotides at the end of C-rich strand and only the G-rich strand can be replicated in its whole. Thus, telomeres continuously erode with each round of replication [40],[43].

Human and mouse telomeres have been shown to erode by about 50-150 bp per cell division, much faster than telomeres from other eukaryotes such as fungi, trypanosomes and flies, which shorten at 3-5 bp per cell division. This suggests that in higher eukaryotes the end replication problem is further exacerbated by an active degradation of the chromosome ends [44]–[46].

Progressive erosion of the telomeric tract sets a finite lifespan to human somatic tissues. After multiple rounds of cell division, telomeres become critically short. Reduced shelterin binding and T-loop stability at telomeres results in critically short telomeres losing their protective function. They become dysfunctional and are recognized as DSBs, leading to the induction of DNA damage signaling pathways . Accumulation of critically short telomeres leads to high occurrence of DDR, which triggers the activation of replicative senescence, which is a state of permanent cell cycle arrest [2],[14],[32],[48].

The suppression of the proliferative capacity of somatic cells is a crucial barrier against cancer development and progression [34]–[36]. The finite replicative life of cells reduces the probability of them acquiring multiple mutations and achieving full progression towards malignancy, thus limiting tumorigenesis. Since cancer cells rely on their immortal potential to maintain aberrant growth, developing mechanisms to counteract telomere shortening and achieve immortality is one of the hallmarks of cancer [4].

1.3.Telomere Maintenance Mechanisms

1.3.1. Telomerase

In most immortal eukaryotic cells, including cancer cells, telomeres compensate for the end replication problem by reactivating the reverse transcriptase telomerase [52]. Telomerase is a specialized ribonucleoprotein complex composed of a highly conserved catalytic subunit with reverse transcriptase activity known as the telomerase reverse transcriptase (TERT) and an associated RNA moiety (telomerase RNA component - TERC), carrying a sequence complementary to the telomeric repeats [43],[53]. Telomerase uses this RNA moiety as a template for reverse transcription to add newly synthesized telomeric repeats to the 3' chromosome ends [54],[55].

Telomerase is expressed in tissues with high turnover, such as hematopoietic cells, skin cells and cells from the gastrointestinal epithelium, to an extent. It is also present at high levels during embryonic development and in many stem cells [12]. Most somatic human tissues do not have sufficient telomerase activity to prevent telomere attrition, so their continued proliferation eventually results in senescence [8]. About 90% of human cancer cells activate telomerase to counter telomere shortening [52].

1.3.2. Alternative Lengthening of Telomeres

A distinct subset of tumors, approximately 10-15%, are telomerase-negative and employ a different mechanism to elongate telomeres, a specialized HDR pathways known as Alternative Lengthening of Telomeres (ALT) [6],[8],[54]. ALT was first described in budding yeast survivors following crisis induced by telomerase inactivation [41]. A few years later, ALT was also characterized in human cells [6],[56].

ALT can occur in common cancers, such as breast carcinomas, but is much more frequent in tumors of mesenchymal or epithelial origin, including, among others, osteosarcomas, liposarcomas, glioblastomas and astrocytomas. [2],[3]. Many of these cancers have particularly poor clinical prognosis.

ALT cancer cell lines maintain many of the common telomeric characteristics such as the duplex TTAGGG repeats, the 3'-overhang, the presence of the shelterin multiprotein complex and the ability to form T-loops. However, in addition to these features, there is a set of characteristics that are specific to the ALT mechanism [57].

Heterogeneous Telomere Length: ALT cells possess a wide range of telomere lengths at different chromosome ends in individual cells, including much longer telomeres than the average for telomerase-positive cells and others so short they are almost undetectable cytogenetically [7].

ALT-Associated PML Bodies: in ALT cells, multiple telomeres cluster in nuclear structures known as ALT-associated PML bodies (APBs). These are ring or disc-shaped aggregates of promyelocytic leukaemia (PML), telomeric DNA and other proteins that are usually bound to the nuclear matrix [7]. They are assembled through multivalent, specialized interaction motifs among PML molecules. They contain the telomere binding proteins, TRF1 and TRF2, as well as a range of proteins involved in DNA recombination and replication. This suggests that APBs provide a ''recombinogenic microenvironment'' to promote ALT at clustered telomere ends [57].

Abundant Extra-Chromosomal telomeric DNA: ALT cells possess a high amount extra-chromosomal telomeric DNA in the form of double-stranded telomeric circles (T-circles), partially single-stranded circles (C- and G-circles) and linear double-stranded DNA [58], [59]. C-circles accumulate exclusively in ALT cells as result of telomeric DNA recombination and C-circle levels correlate with the levels of telomere DNA synthesis in ALT cells. Indeed, C-circle abundance is widely used as a marker for ALT activation and the C-circle detection assay represents the most robust and quantitative ALT assay so far [60],[61]. How C-circles are generated during ALT remains elusive, but they may represent the product of excision of over-elongated telomeres

T-SCEs were found to occur more frequently in ALT cells than in telomerase-positive cell lines or healthy cells, without an increase in sister chromatid exchanges frequency elsewhere in the genome [8],[13],[62].

Elevated levels of the telomeric long noncoding RNA TERRA: Northern blot experiments with total RNA from cell cultures has shown that TERRA is more abundant in ALT cell lines than in telomerase-positive cells [13],[14].

ATRX/DAXX Inactivation: inactivation of one or both of the ATP-dependent chromatin remodelers alpha-thalassemia/mental retardation X-linked (ATRX) and death domain-associated protein-6 (DAXX) is commonly found in ALT cancers [9],[13]. ATRX and DAXX form a complex that deposits the histone variant H3.3 at heterochromatic loci, including telomeres. Reduced ATRX and/or DAXX activity may be the reason for the altered chromatin state of ALT telomeres and possibly the deregulation in TERRA transcription and T-SCEs.

The HDR ALT pathway relies on break-induced replication (BIR) [63], a conservative DNA synthesis-based repair process that is engaged at one-ended DNA DSBs and arrested replication forks. BIR results in homology-directed synthesis of new telomeric DNA using an existing telomeric sequence from an independent chromosomal telomere as a copy template [11]. There are two types of BIR, either RAD51- or RAD52-dependent, which were firstly identified in ALT yeasts [64], [65]. In human ALT cells, BIR is exclusively RAD52-dependent and also relies on the DNA polymerase δ accessory subunits POLD3 and POLD4 and on proliferating cell nuclear antigen (PCNA) [57]. New evidence has suggested that there might exist also a RAD52-independent ALT pathway in humans [11]. Since ALT is dependent on HDR of damaged DNA, this creates a situation in which at least a subset of telomeres in ALT cells is experiencing physiological damage at all times [66]. This sustained damage derives from an ALT-specific telomeric replication stress (ATRS)

The triggers of ATRS are not fully known, but many hypothesis can be proposed. Firstly, telomeres are inherently difficult regions to replicate due their sequence repetitive nature and to the tight association of telomeric DNA with heterochromatin marks and telomeric proteins. Telomeres also have an abundance of higher order

telomeric structures that may contribute to ATRS. These structures include the previously described T-loops (See Section 1.2.1) and telomeric R-loops (tel-Rloops), formed by annealing of TERRA with the C-rich strand of the telomere [67] (Further explained in Section 1.4.2). Finally, G-quadruplexes can form when the G-rich telomeric strand exists in ss state, for example, at the displacement loop of a T-loop or a telR-loop [68].

To counterbalance the effect of replication stress triggers, which could activate DNA damage checkpoints that halt cell proliferation, alleviators of ATRS are constantly active in ALT cells. The endoribonuclease RNaseH1 is a major alleviator. It associates with ALT telomeres, where it degrades the RNA moiety of telR-loops. Depletion of RNaseH1 in ALT cells has been shown to increase telR-loops, ATRS and C-circles and to causes rapid loss of entire telomeric tracts [14]. Another ATRS alleviator is the ATP-dependent DNA annealing helicase SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A-Like 1 (SMARCAL1). SMARCAL1 is able to restart arrested replication forks through fork regression, is known to be enriched at ALT telomeres and its depletion using siRNAs augments ATRS, telomeric DNA damage and c-circles at telomeres [69]. Lastly, ALT is alleviated by Fanconi anemia, complementation group M (FANCM), a highly conserved protein, with ATPase and DNA translocase activity, belonging to the Fanconi anemia (FA) disorder core complex. Recent reports revealed that FANCM limits telomeric replication stress and damage and, in turn, ALT activity by suppressing accumulation of telomeric R-loops and by regulating the action of the BLM helicase at telomeres. As a consequence, FANCM depletion in ALT cells leads to exaggerated ALT activity and ultimately cell death [67],[71],[72]

Accumulation of telR-loops, G-quadruplexes at telomeres interfere with DNA replication and subsequently lead to the collapse of replication forks and formation of one-ended DSBs [11]. Replication stress also causes the enrichment of telomere proteins with the interaction motifs for PML aggregation, which trigger APB formation. Telomere ends cluster at APBS, which are enriched with proteins responsible for DNA repair, recombination and replication. In human ALT cells, telomere ends are elongated in APBS through RAD52 and POLD3/POLD4-dependent BIR [11]. The framework of the ALT pathway in humans is presented in Figure 3.

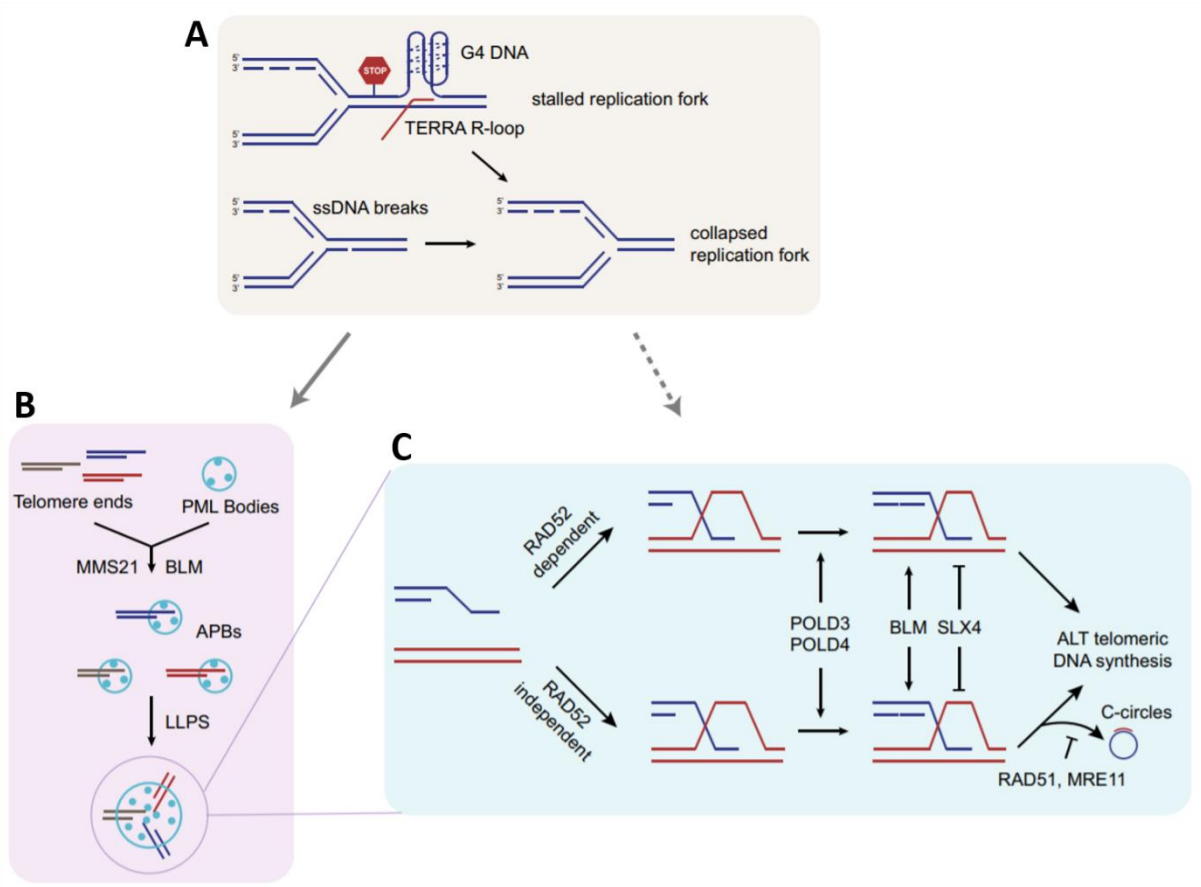

Figure 3: Framework of the ALT pathways. (A) Replication stress at telomeres leads to accumulation of R-loops, G-quadruplexes and DNA single-strand breaks that interfere with DNA replication, leading to collapse of replication forks and formation of one-ended DSBs. (B) the replication stress or DNA damage at telomeres enriches telomere proteins with PML binding motifs, which recruit PML and trigger APB formation. The clustering of telomeres and enrichment of DNA repair, recombination and replication proteins in APBs drive ALT. (C) In APBs, BIR is triggered by the one-ended DSBs at telomeres. ALT can take place through RAD52-dependent and -independent BIR pathways. The conservative DNA replication during BIR is dependent on POLD3/POLD4, promoted by BLM and inhibited by SLX4. C-circles are generated by the RAD52-independent BIR pathway, which is suppressed by RAD51 and MRE11. Adapted from [11].

1.4.TERRA

For many years it was believed that telomeres were transcriptionally silent [47], due to their heterochromatic and gene-less nature and their ability to silence the transcription of experimentally inserted subtelomeric reporter genes (telomere position effect) [72]. This long lasting belief was eventually overturned with the discovery of transcriptional activity at telomeres of protozoa [73].

Over the following years, similar transcriptional activity was observed at chromosome ends of many species and telomeres are now proven to be transcribed by RNA polymerase II (RNAPII), giving rise to a class of long noncoding RNAs containing telomeric repeats called TERRA (Telomeric Repeat-containing RNA) in multiple species including humans [12],[74], mice [74], zebrafish, yeast and plants [75]. In humans, TERRA was first

detected by northern blot analysis of RNA from a human cervical cancer cell line (HeLa) using strand-specific telomeric and subtelomeric probes [12].

During the past decade it has become clear that, via its interactions with several protein partners and its ability to form RNA:DNA hybrids with telomeric repeat-containing sequences, TERRA is an active participant of several mechanisms occurring at telomeres, including formation of heterochromatin, proper capping of chromosome ends, regulation of telomerase and telomere elongation by ALT mechanisms [75].

1.4.1. TERRA Biogenesis and Characterization

TERRA molecules are transcribed by RNAPII from the subtelomeric regions toward the chromosome ends, with the C-rich telomeric strand being used as template (Figure 4),[12]. Thus, TERRA molecules comprise chromosome-specific subtelomeric sequences at their 5' end and terminate with tandem arrays of G-rich telomeric repats (5'-UUAGGG-3'). The subtelomeric-derived sequence can be amplified by RT-qPCR [12], which is the basis for many TERRA expression analysis methods and can be used to distinguish the chromosome of origin of a TERRA transcript although some chromosomes have highly similar subtelomeric sequences [12], [16],[76]. These qPCR studies have consistently supported the telomeric origin of human TERRA and found TERRA transcription to occur from subtelomeric promoters located on at least two-thirds of chromosomes [75],[38],[40]. TERRA transcripts range in size from about 100 bases up to more than 9 kilobases and are mostly detected in nuclear cellular fractions where they colocalize with telomeres [22],[28].

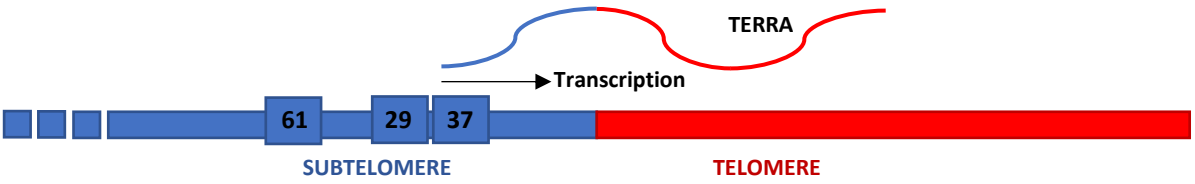

Figure 4: Schematic representation of TERRA generated by transcription of eukaryotic chromosome ends. RNA and DNA sequences of telomeric or subtelomeric origin are indicated in red or blue, respectively. Black arrow indicates the direction of transcription. 61-29-37 repeats TERRA promoter is also indicated, within which is the TERRA transcription start site.

In humans, most TERRA transcription start sites are within CpG islands, high CG content regions located within a subset of subtelomeres, proximal to the telomeric repeat tract (Figure 5B) [16]. The first human subtelomeric promoter to be characterized within CpG islands is a repetitive region conserved across telomeres, located directly upstream of TERRA transcription start site and about 1 kb away from the telomeric tract. It is known as the "61-29-37 bp repeats" because it comprises three different repetitive DNA tracts. First, a most centromere-proximal tract with a variable number tandem 61- base-pair-repeats, followed by tandem 29-bp tandem repeats and finally, a most distal set of tandem 37-bp repeats [16][77] (Figure 4). The 61-29-37 repeats have been identified in at least 20 human chromosome ends by bioinformatics analysis combined with in situ hybridization experiments [78] (Figure 5A). The 29-37 bp sequences, specifically, have been proved to be bound by active

RNAPII in vivo and to support transcription in promoter reporter assays in cells [16]. On the other hand, the 61 bp repeat does not have promoter activity, being an insulator and a binding site for the chromatin organizing factor CTCF (CCCTC-binding factor), which acts as a key regulator of TERRA expression [79].

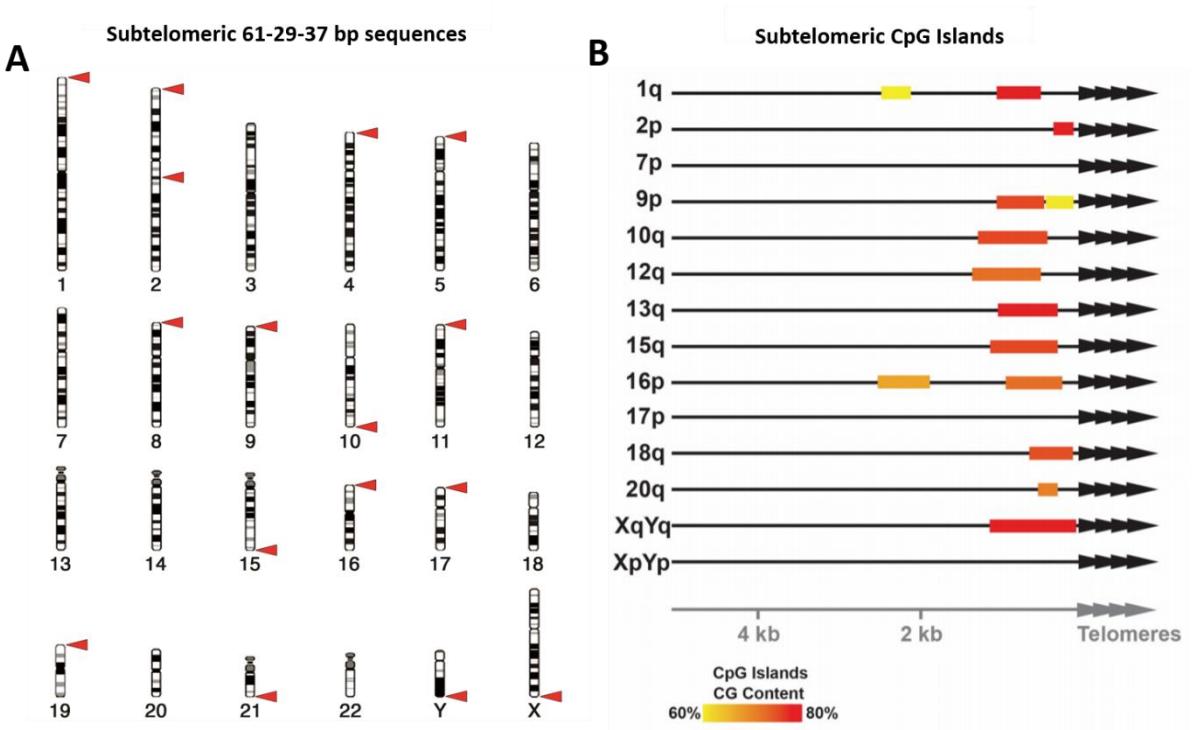

Figure 5: (A) Chromosomal distribution of 61-29-37 repeat sequences. Red arrows indicate chromosomal loci containing 61-29-37 repeats as derived from BLAST analysis. Adapted from [16]. *(B) Representation of position and CG content of subtelomeric CpG islands in the indicated chromosome ends. Adapted from* [18].

Approximately 10% of TERRA species are polyadenylated at their 3' ends in humans. This polyadenylation influences TERRA stability and association with chromatin. Human TERRA 5' ends are also capped by 7-methyl-guanosine cap structures [51]. TERRA transcription is regulated by the heterochromatic state of telomeres and subtelomeric regions. Methylation of telomeric DNA as well as lysine 9 of histone h3 (H3K9) lysine 20 of histone h4 (H4K20) trimethylation repress TERRA expression [75], while histone acetylation promotes TERRA transcription [47]. The evidence of active TERRA transcription from subtelomeric promoters and regulation through various factors supports the idea that TERRA plays a fundamental role in telomere biology.

It is documented that telomere length regulates TERRA expression levels. Short telomeres are associated with higher TERRA levels and an increased transcriptional activity that correlates with a reduced density of repressive histone marks at telomeric repeats [80]. It has also been demonstrated that TERRA transcripts associate with chromosome ends. Localization of TERRA at chromosome ends is increased with the depletion of components of the nonsense mediated RNA decay (NMD) pathway and with the depletion of members of the heterogeneous nuclear ribonucleoprotein family, which bind TERRA [12]. This suggests that TERRA molecules are actively displaced from telomeres. They may be recruited at chromosome ends through interaction with stable constituents of the telomeric structure, including the shelterin proteins TRF1 and TRF2 [81]. TERRA has been

found to bind to other proteins which also localize at telomeres, such as the heterochromatin protein 1 (HP1) and MORF4L2, a component of the NuA2 histone acetyltransferase complex [15]. These findings evidence the complex connection of TERRA to telomeres.

1.4.2. TERRA and Telomere Maintenance

TERRA transcripts interact directly with telomeres forming telR-loops (Figure 6). TelR-loops result from bonds formed between G-rich (UUAGGG)n TERRA transcripts and C-rich (CCCTAA)n complementary DNA strands. When a telR-loop is formed, the two strands of dsDNA are separated, with one binding to the complementary TERRA RNA and the other being displaced [82]. They are a major trigger of ATRS, as mentioned in Section 1.3.2, as they can physically interfere with replication fork progression and prime the initiation of BIR. Therefore, they play a fundamental role in ALT mechanisms [14].

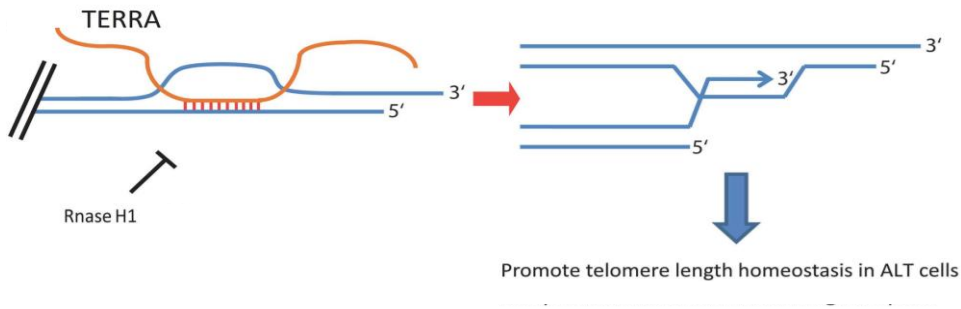

Figure 6: Proposed role of TERRA in telomere length homeostasis of ALT cells. TERRA molecules form DNA:RNA hybrids, or telR-loops, at telomeres. telR-loop formation is inhibited by RNase H1 in ALT cancer cells. Telomeric R-loops promote homologous recombination among telomeres, which maintains telomere length homeostasis in ALT cancer cells. Adapted from [25]

TelR-loops have been found in various human cancer cells, both telomerase-positive and ALT. Formation of telR-loops at chromosome ends is influenced by the sequence characteristics of the region adjacent to the telomeric tract, including the TERRA promoter and TSS and is promoted by elevated levels of TERRA and/or by perturbation of the normal chromatin structure in these regions [83].

While R-loops occur spontaneously, they also pose a threat to genome integrity as their formation associates with mutations, recombination, replication fork stalling and chromosome rearrangements. For this reason, telR-loop formation is tightly controlled within the cell and hazardous telR-loop structures are removed by the activities of different enzymes such as RNase H1 and RNase H2 which degrade the RNA part of a DNA-RNA hybrid [84].

Persistent telR-loops might trigger DNA damage and subsequent BIR, which are the cornerstones of ALT, as DNA DSBs arise from a collision between the replication machinery and a telR-loop [84]. Hence, TERRA is believed to play a pivotal role in sustaining DNA replication in ALT through R-loop formation.

TERRA is a potential target for therapeutic approaches to ALT cancers. Considering the role of TERRA as an activator in ALT, it is tempting to consider therapeutic approaches consisting in suppression of TERRA in order to block the ALT mechanism in cancer cell populations, and lead to activation of senescence. However, an arguably more effective potential treatment could consist of upregulating TERRA targeting appropriate transcription factors. As mentioned before, ALT relies on maintaining a very delicate balance of ATRS, enough to trigger HDR but not so much that could result in genomic instability. Creating excessively high levels of TERRA in a cell population could disrupt this balance, causing an excess of telR-loops and elevating ATRS levels to a point that would cause widespread genomic instability and apoptosis in ALT cancer cells

1.4.3. Studies on TERRA Transcription Profiling

Due to the established role of TERRA in telomere maintenance, there has been considerable interest in studying TERRA transcriptional profiles in human cells. This has been a controversial issue, with conflicting reports from different laboratories.

In the report that first characterized TERRA CpG promoters, Nergadze et al. used rapid amplification of complementary DNA (cDNA) ends (RACE) to identify the 61-29-37 repeats with TERRA promoter activity, conserved across a series of subtelomeres (15q, 16p and 19p and Xq/Yq), suggesting that TERRA transcription occurs at several chromosome ends [16].

Since then, multiple studies have used chromosome specific RT-qPCR to study TERRA expression across a broad panel of chromosome ends in multiple human cell lines, with their findings confirming that TERRA is transcribed from multiple telomeres [44],[48].

Porro et al. performed Illumina RNA-seq experiments using TERRA-enriched RNA from HeLa cells in 2014 and found transcription initiation of TERRA occurring at several chromosome ends. They also found substantial variation of transcription start site distance from telomeric repeat tracts among different telomeres, extending in some cases for several kilobases into the subtelomeric regions. The authors identified 10 distinct chromosome ends where transcription into TERRA molecules started as far as 5–10 kb upstream of the subtelomere, while in eight other chromosome ends, the RNA-seq data suggested that transcription starts in close proximity to the TTAGGG-tracts [15].

Montero et al., contrarily, reported that TERRA stems mainly from chromosome end 20q, in 2016 [17]. The authors found that deletion of the 20q locus in U2OS cells using the CRISPR-Cas9 technology causes a dramatic decrease in total TERRA levels, suggesting that TERRA transcripts expressed from this telomere represent a substantial fraction of the total TERRA population and 20q-TERRA ablation leads to dramatic loss of telomeric sequences and the induction of a massive DNA damage response. These findings suggest that TERRA-20q transcripts act *in trans* by participating in telomere protection at multiple chromosome ends, in U2OS cells.

The first study of TERRA transcription which measured the number of TERRA copies per cell in human cell populations, comes from Feretzaki et al. in 2019, where the authors quantified TERRA molecules from different chromosome ends in a variety of human cell lines, using a standard curve for absolute quantification of RT-qPCR TERRA measurements results [18]. The study demonstrated that TERRA is expressed in variable amounts from a large set of chromosome ends.

A better understanding of the dynamics of TERRA molecules, for example, by characterizing whether TERRA transcripts act only *in cis*, at the transcribing telomere, or also *in trans*, by relocating to other chromosome ends is still necessary [79]. Beishline et al. suggested that TERRA transcripts act to facilitate DNA replication *in cis*, at their telomere of origin, driven by CTCF [86]. CTCF-binding site mutation and decreased levels of TERRA transcription directly impacted the DNA replication efficiency of the telomere from which it is transcribed, indicating that it has a direct *cis*-acting effect on telomere stability. However, in a more recent report, Feretzaki et al. revealed a RAD51-dependent pathway that provides a mechanism for TERRA to be recruited to independent telomeric loci *in trans*, where it mediates *in trans* R-loop formation [87].

Understanding TERRA transcription dynamics also has a clear potential impact on clinical practice. For instance, identification of chromosome ends that transcribe high amounts of TERRA acting in trans in other chromosome ends could guide the development of therapeutical or diagnosis practices targeting specific chromosome ends.

1.5. Methods of TERRA Quantification

1.5.1. Northern Blot

In the northern blot assay, a sample of RNA is separated by denaturing agarose gel electrophoresis, transferred to a membrane and immobilized. A radiolabeled C-rich telomeric probe is then used to detect TERRA transcripts. Northern blot detects total TERRA transcripts in a sample, while techniques like RT-qPCR are dependent on chromosome end specific primers. Although northern blotting is able to detect a global increase in TERRA transcripts, it is ineffective in measuring minor changes in transcription or quantifying TERRA levels from individual chromosome ends [76].

1.5.2. Oxford Nanopore Technologies Sequencing (ONT-seq)

Oxford Nanopore Sequencing is a technology that enables direct, real-time analysis of long DNA or RNA fragments. It functions by monitoring changes into an electrical current as nucleic acids are passed through a protein nanopore, inserted in a synthetic membrane. Single molecules entering the nanopore cause characteristic disruptions in the current. The resulting signal is decoded to provide the specific DNA or RNA

sequence. This technique can be used in transcriptomic to estimate the abundance of a target transcript, such as TERRA [88].

Recently, Oxford Nanopore Technologies (ONT) launched a portable sequencer called MinION, which offers the possibility of sequencing long RNA molecules after reverse transcription [89]. The Oxford Nanopore Technologies Sequencing (ONT-seq) assay consists of reverse transcribing an RNA sample and driving the samples through the nanopore membrane electrophoretically. MinION nanopore returns the full transcriptome of the sample separated into reads (i. e. inferred sequences of base pairs corresponding to all, or part of, a single cDNA fragment). The device can return long reads that individually span the full length of a single transcript. This generates a read library of the sample. The reads are then aligned back to a reference genomes. Transcripts that are more abundant generate a higher amount of reads. When reads are mapped to the genome, transcript abundance is estimated by quantifying the read coverage of the genome sequence from which transcription occurs [88], [89].

1.5.3. RT-qPCR

The central technique used in this project for TERRA quantification is quantitative reverse transcription polymerase chain reaction (RT-qPCR). It is one of the most sensitive and reliable quantitative methods for gene expression analysis and is a widely used tool for nucleic acid quantification in biological fields [90],[91].

The principle of PCR amplification is controlled amplification of double-stranded DNA sequences. A reaction is prepared in which a polymerase enzyme amplifies a short specific part of the template DNA (amplicon) in cycles

This amplification has three distinct phases [91]: the first is the exponential phase, in which PCR product increases exponentially since the reagents are not limiting. It is followed by the linear phase, in which product increases linearly as PCR reagents become limiting. Lastly, PCR reaches the plateau phase, during which the amount of product will not change because some reagents become depleted. Real-time PCR exploits the fact that the quantity of the PCR DNA product in exponential phase is proportional to the quantity of initial template, under ideal conditions[90], [92].

During the exponential phase, the PCR product is doubled with each cycle if efficiency is perfect (100% efficiency). Making the PCR amplification efficiency close to 100% in the exponential phase if the PCR conditions depends on a series of factors such as primer characteristics, template purity and amplicon length [91].

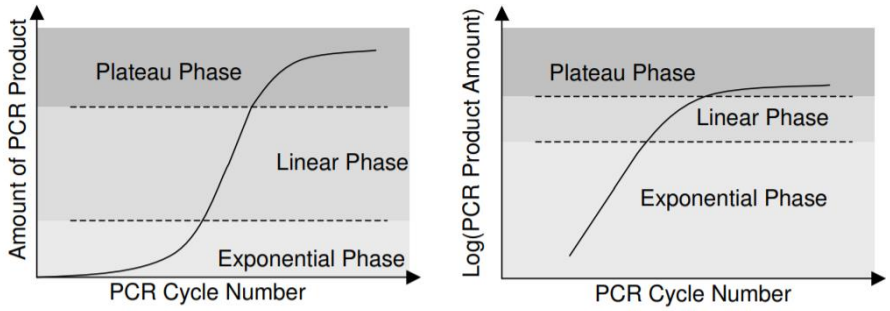

Figure 7: (A) Theoretical plot of PCR cycle number against PCR product amount. Three phases can be observed for PCRs: exponential phase, linear phase and plateau phase. (B) Theoretical plot of PCR cycle number against logarithm PCR product amount. Adapted from [91]

The dynamics of the amplification reaction can be measured through DNA binding dyes like SYBR green or DNA hybridization probes, since there is a positive correlation between fluorescent dye detection and the number of amplified copies of template (amplicons) present in the reaction.

This correlation is measured by plotting of logarithm 2-based transformed fluorescence signal versus cycle number. This yields a linear range at which logarithm of fluorescence signal correlates with the original template amount. At this point a threshold is set for analysis. Fluorescence is monitored during the entire PCR process. The more copies of the DNA target amplicon the sample contains, the less cycles it will take for fluorescence to reach the threshold. The cycle number at the threshold level of log-based fluorescence is called a Ct (cycle threshold) value. This value is the primary statistical metric of interest in the RT-qPCR assay [91].

There are two ways of quantifying RT-qPCR products, absolutely and relatively, both of which are used in this project.

Absolute quantification makes use of a standard curve. The curve is made from RT-qPCR reactions in a serial dilutions series of a standard with known copy number and returns the corresponding CT values. Thus, the copy number of the target gene in the qPCR input can be derived from the CT value of the target gene. It can determine absolute transcript copy number, making it essential for the purposes of this project [91], [92]. Standard curves are highly reproducible and allow the generation of highly specific, sensitive and reproducible data. However, the standard curve itself needs to be thoroughly validated, as the accuracy of absolute quantification depends entirely on the accuracy of the standards.

Relative quantification relies on the comparison between expressions of a target gene versus an endogenous reference gene. [7]. Target quantification consists in determining the fold-change of the target gene in relation to the reference gene, across every sample. The most used relative quantification method is the ΔCt method, which is also used in this project.

For this method, the differences (Δ) in the threshold cycle (Ct) between the target gene and the endogenous reference gene are measured, determining the ΔCT value for each sample. Reference genes are usually highly expressed genes, whose expression does not change between samples of the same cell line.[93].

$$\Delta\mathrm{CT\ (sample)} = \mathrm{CT\ target\ gene} - \mathrm{CT\ reference\ gene}$$

Assuming that the PCR efficiencies of the target gene and endogenous reference gene are comparable, the normalized level of target gene expression is calculated by using the formula [93]:

$$\text{Normalized target gene expression level in sample} = 2^{-\Delta C}$$

The use of RT-qPCR for reverse transcribed cDNA from TERRA RNA has the ability, contrary to classical analysis of TERRA by northern blotting, to accurate quantify TERRA molecules expressed from individual chromosome ends through the analysis of a chromosome end-specific amplicon. It is, thus, an ideal method for studying TERRA transcription loci and effects of regulatory factors on TERRA transcription, in a chromosome end-specific way.

1.6. Thesis Background and Objectives

1.6.1. Background: ONT-seq Experiment

To obtain TERRA expression profiles for every human chromosome end, the Azzalin lab at iMM performed Oxford Nanopore long-read sequencing (ONT-seq) of a cDNA library produced from purified TERRA RNA from two ALT (U2OS, SAOS-2) and two telomerase-positive (HEK293T, HeLa) cell lines. Normalization was performed by dividing TERRA reads per total reads in the RNA sample. The reads were mapped to a publicly available assembly of human subtelomeres. Read density profiles were generated by adding the number of reads aligned at each position along the subtelomeric regions. Read coverage of each individual chromosome end allowed for estimation of TERRA expression. TERRA expression profiles were obtained for every human chromosome end, a subsection of which is presented in Figure 8.

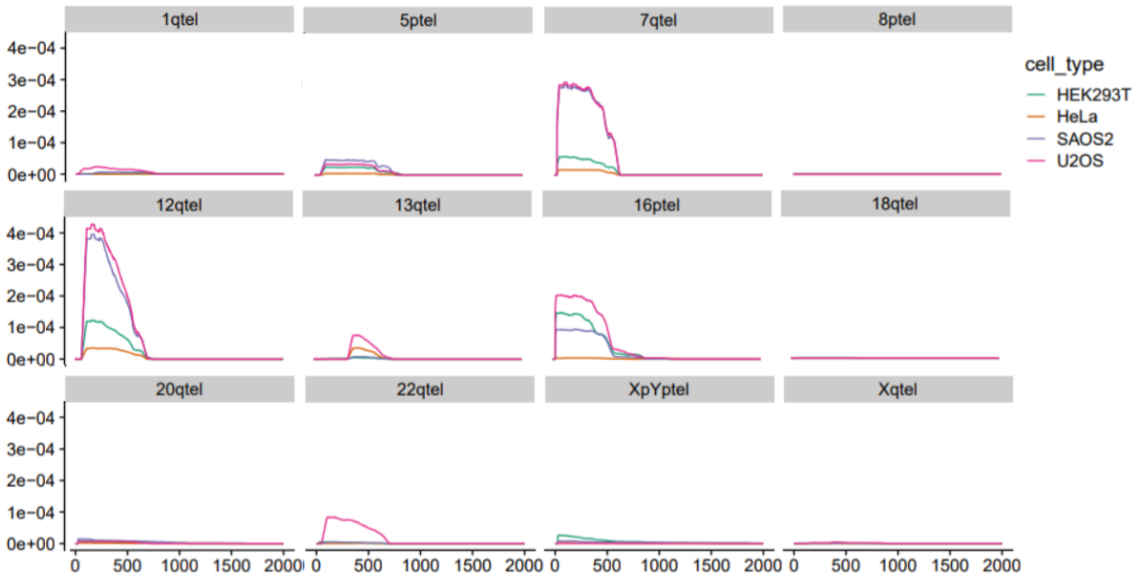

Figure 8: Oxford ONT-seq of TERRA expression for a selection of chromosome ends. x axis displays the 2000 bp after the telomeric repeats in each subtelomere, with numbering beginning from 0 immediately after the telomeric repeats. y axis is TERRA long-read coverage, on a log scale.

The ONT-seq also allowed the identification of putative TERRA transcription start sites (TSS) in each chromosome end. TSSs are the locations of the majority of 5' ends of mapped TERRA reads in the chromosome end. Putative TERRA TSSs are usually found immediately upstream of high read coverage areas.

The ONT-seq data demonstrated low TERRA expression across most chromosomes ends (Figure 8) in telomerase-positive cell lines, consistent with the literature. It also suggested highly unequal distribution of TERRA expression between telomeres in ALT cells. Two chromosome ends, the long arms of chromosomes 7 and 12 (7q and 12q chromosome ends) emerge as producing the most abundant TERRA transcripts in ALT cells. The 7q and 12q chromosome ends contain a subtelomeric CpG island proximal to the telomeric repeats that is highly conserved between the two ends. Comparison of the two subtelomere sequences reveals only a few individual polymorphisms between them (Figure 9) and a 30 bp sequence present at 12q subtelomere that is missing from the 7q subtelomere. This conserved sequence will henceforth be called the 7q_12q subtelomeric sequence. It is interesting to note that the 7q and 12q chromosome ends do not contain the previously characterized 61-29-37 TERRA promoter. This indicates that the ONT-seq results could be unveiling new TERRA promoters within subtelomeric CpG islands.

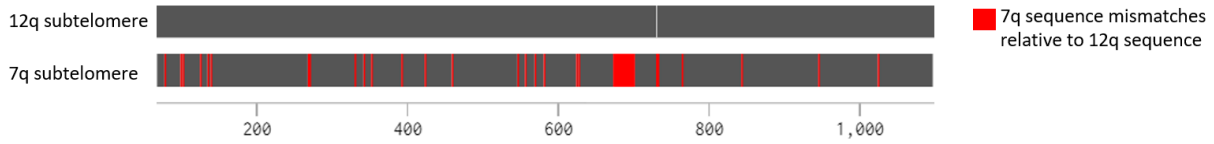

Figure 9: BLAST of the first 1000 bp of the 7q subtelomere immediately following the telomeric repeats, to the same region of the 12q subtelomere. x axis indicates bp distance from the telomeric repeats. Red bars indicate polymorphisms between the two sequences. Representation made with benching software (https://benchling.com/)

The high levels of TERRA transcribed from 7q and 12q compared to other ends in ALT cells and to the same end in telomerase-positive cells, makes these telomeres compelling targets for clinical purposes. The Azzalin lab hypothesized that quantification of 7q- and 12q-TERRA might represent a valuable and sensitive tool for ALT diagnosis. The only currently available diagnosis assay is the C-circle assay, which not only has ALT quantifying capabilities, but responds quickly to changes in ALT activity [61]. However, RT-qPCR quantification of chromosome end-specific TERRA could be an excellent complement to the c-circle assay. It also has quantitative capabilities and, possibly, the same ability to quickly respond to fluctuations in ALT activity.

1.6.2. Goal 1: Validation of TERRA quantification as a new ALT diagnosis tool

The TERRA ONT-seq results served as the launching pad of this project. With 7q and 12q chromosome ends suggested as significant players in TERRA expression, the first goal of this project was to validate TERRA quantification as a new diagnosis tool by measuring and comparing the levels of 7q- and 12q-TERRA in several ALT and telomerase-positive cancer cell lines.

For this diagnosis procedure to be viable, the difference in 7q and 12q TERRA expression should be of orders of magnitude between ALT and telomerase-positive cancer cell lines, to allow for the establishment of diagnosis threshold that ensures reliability and sensitivity, with little risk of false positives or negatives due to situational variations in either the technical procedure or in the RNA sample.

1.6.3. Goal 2: Validation of ONT-seq Results through RT-qPCR

The second stage of this project moved beyond the focus of quantifying TERRA expression with a potential diagnostic goal, into a more extensive validation of the ONT-seq results through RT-qPCR. A panel of 4 cell lines was studied, the same cell lines used in the ONT-seq experiment, ALT U2OS and SAOS2 osteosarcoma and telomerase-positive HeLa cervical cancer and HEK embryonic kidney cells.

1.6.4. Goal 3: Identification and characterization of the 7q and 12q TERRA Promoter

Chromosome ends 7q and 12q display high TERRA expression in ONT-seq data despite not containing the previously characterized 61-29-37 repeats promoter. This drove the third part the project, the identification and characterization of a new TERRA promoter sequence present within the CpG islands of 7q and 12q telomere ends.

Finding new sites of active TERRA transcription could be of great use to better understand the mechanics and TERRA transcription and functions and to guide clinical practices. A new TERRA promoter in these sequences would add to the current catalogue of known TERRA promoters and contribute to a better mapping of the TERRA transcription profile in human cancer cells.

2. Materials and Methods

2.1.Cell lines and Culture Conditions

The ALT cell lines used in this project were U2OS and SAOS2 osteosarcoma cells. The telomerase-positive cell lines used were HOS osteosarcoma, HeLa cervical carcinoma, HEK293T embryonic kidney, and HCT116 and HCT 116 DKO colorectal carcinoma cells. U2OS, HeLa and HEK cells were cultured in high glucose DMEM, GlutaMAX (Thermo Fisher Scientific) supplemented with 5% tetracycline-free fetal bovine serum (Pan BioTech) and 100 U/ml penicillin-streptomycin (Thermo Fisher Scientific). HCT116 and HCT116 DKO cells were cultured in the same conditions except supplemented with 10% fetal bovine serum. SAOS2 and HOS cells were cultured in high glucose DMEM/F12, GlutaMAX (Thermo Fisher Scientific), supplemented with 5% tetracycline-free fetal bovine serum (Pan BioTech), 100 U/ml penicillin-streptomycin (Thermo Fisher Scientific) and nonessential amino acids (Thermo Fisher Scientific). The cells were maintained in 5% CO_2 incubator at 37 ºC.

2.2.RNA Preparation

2.2.1. Extraction

RNA from cultured cells was isolated using NucleoSpin RNA (Macherey-Nagel). To begin, cells cultured in 15 cm plates were harvested by incubating with 2mL 0,05% trypsin-EDTA (ThermoFisher) for 2 min, resuspended in DMEM media and counted using a hemocytometer (Neubauer). Cells were then pelleted by centrifugation at 500 xg at 4°C. Supernatant was removed and cells were resuspended in 1x PBS for washing, then pelleted once more. Supernatant was removed and cells were resuspended in 350 µL Lysis Buffer RA1 from Macherey-Nagel Kit.

Following this, 3.5 µL ß-mercaptoethanol was added and the mix was vigorously vortexed. Samples were filtrated through NucleoSpin® Filter by centrifuging for 1 min at 11000 xg to reduce viscosity and clear the lysate. 350 µL ethanol (70 %) was added to the homogenized lysate and mix by pipetting up and down. The lysate was loaded onto a NucleoSpin® RNA Column placed in a Collection Tube and centrifuged for 30 s at 11000 x g. 350 µL of Membrane Desalting Buffer was added and the lysate was centrifuged at 11000 xg for 1 min to dry the membrane.

95 µL of DNase reaction mixture containing 90% reconstituted rDNase was added directly onto the center of the silica membrane of the column and samples were incubated at room temperature for 15 min.

The silica membrane was washed by centrifuging by centrifuging for 30 s 11000 xg with 200 µL Buffer RAW2, then with 600 µL Buffer RA3 and finally for 2 min with 250 µL Buffer RA3. RNAs were eluted in a 1.5 ml Eppendorf tube in 176 µL RNase-free H2O, by centrifuging at 11000 xg for 1 min. RNA concentrations were measured on a NanoDrop spectrophotometer.

The number of cells isolated for RNA extraction for each RT-qPCR assay is summarized in the table below (Table 1).

Table 1: Summary of the number of cells isolated for RNA extraction for each assay

RT-qPCR Assay	Cell Line	Cells isolated for RNA Extraction
Results from 3.1	U2OS	5 to 10 million
	HOS	
Results from 3.2	U2OS	4 to 5 million
	HeLa	
	SAOS2	
	HEK	
Results from 3.3	HCT116	4 million
	HCT116 DKO	

2.2.2. In-Solution DNase Treatments

Two in-solution DNase digestions were performed to remove any trace of genomic DNA contaminants. Eluted RNA was mixed with 4 µl Qiagen DNase I (5 Kunitz units/µl) and 20 µl 10X RDD buffer in 200 µl total volume and incubated for 1 hour at room temperature. Subsequently, 40 µl (1/10 volume) NaOAC 3 M (pH=5,2) and 200 µl RNase-free H2O were added and samples were mixed by inverting 8 times. 1100 µl (3 volumes) ethanol (100%) was then added, samples were mixed by inverting 8 times again and RNA was precipitated at -80 ºC for at least 1 hour. Samples were centrifuged at 13,000 rpm for 20 min at 4ºC, then washed with 800 µl ethanol (80%) and centrifuged again with the same parameters, supernatants were removed and RNAs were air-dried for 10 min at room temperature. RNAs were re-hydrated in H_2O on ice for 45 min, flicking the tube every 10 min. The DNase treatment and precipitation process were repeated and the samples were finally re-hydrated in 82 µL RNase-free H_2O. RNA concentrations were measured on a NanoDrop spectrophotometer. RNA was stored at -80 ºC.

2.3.Genomic DNA Preparation

Cells were harvested and pelleted from 10cm culture plates, as described in section 2.2.1 and genomic DNA (gDNA) was isolated using Wizard® Genomic DNA Purification Kit (Promega). The cell pellets were resuspended in 600 µl of Nuclei Lysis Solution and repeatedly pipetted up and down to lyse cells. 200 µl of Protein Precipitation Solution were added and samples were vortexed vigorously at high speed for 20 seconds. Samples were chilled on ice for 5 min, then centrifuged for 4 min at 13000 xg, forming a tight white pellet of precipitated protein. The

supernatants containing the DNA were removed and transferred to a clean 1.5ml microcentrifuge tube containing 600 µl of room temperature isopropanol. The solutions were mixed by inversion until the white thread-like strands of DNA formed a visible mass. The samples were then centrifuged for 1 min at 13000 xg at room temperature. The supernatants were carefully decanted. 600 µl of room temperature ethanol (70%) was added and the tubes were inverted several times to wash the DNA. DNAs were centrifuged for 1 minute at 13000 xg at room temperature. Ethanol was aspirated and the pellets were air-dried for 10–15 min. DNAs were rehydrated in H2O by incubation at 65°C for 1 hour, or overnight at 4°C. DNA was stored at 4°C.

2.4.Reverse Transcription

2.4.1. Reverse Trancription Primer Design

Two distinct RT primers were designed for the reverse transcription step (Table 2)(Figure 10). Primer o1 consisted of five repeats of the mammalian C-rich telomeric sequence 5'-(CCCTAA)-3', complementary to the 5'-UUAGGG-3' repeats at the TERRA transcript 3' end, so that cDNA synthesis was initiated from the TERRA 3' and covered the entire subtelomeric-derived part of the transcript.

Table 2: Reverse Transcription Primers used in this study

Number	Primer	Sequence (5' -> 3´)
o1	C-rich RT primer	CCCTAACCCTAACCCTAACCCTAACCCTAA
o2	7q_12q subtelomeric RT primer	GCACGGCAGTGTGCGGATTG

Primer o2 was designed to be complementary to a section the TERRA subtelomeric-derived sequence proximal to the telomeric-derived repeats, which was known from the ONT-seq results. As before, synthesis occurred in the TERRA 5' direction and covered the region targeted for amplification by qPCR.

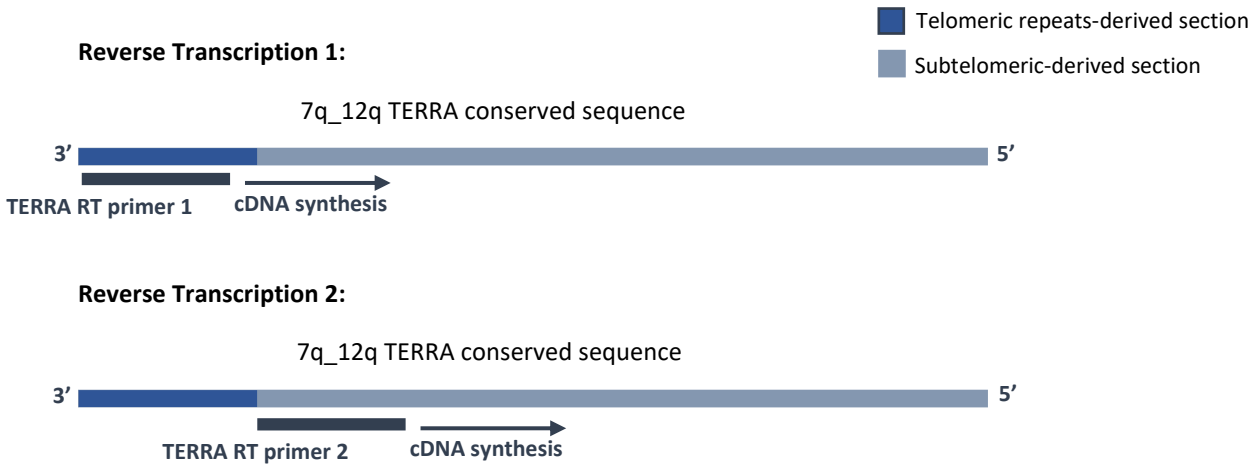

Figure 10: Representation of the reverse transcription reaction of the 7q_12q TERRA sequence. Telomeric and Subtelomeric derived sequences represented in dark blue or light blue, respectively. The two reverse transcription primers used in the study are represented under their complementary regions in dark blue (bar size does not reflect primer length). Arrows indicate cDNA synthesis direction.

2.4.2. cDNA Synthesis

For each RNA sample, 4 µg (in the experiment described in Section 3.1) or 5 µg (in the remaining experiments) were reverse transcribed, side by side with a control, to each no reverse transcriptase was added (-RT internal control). RT-qPCR assays were based on protocols by the Lingner lab [18], [76] and on protocols from Azzalin lab.

The following components were added to a nuclease-free 200 µl PCR tube; 1 µl reverse transcription primer (10 µM), 1 µl endogenous control gene RT primer (β-actin) (1 µM), 1 µl dNTP mix (10 mM) and 4 to 5 µg of one of the samples of total RNA. The mixture was heated for 3 min at 85 ºC, then passed on ice for 2 min. To each PCR tube. 4 µl 5X First-Strand buffer and 2 µl DTT (0.1 M) were added. The mixture was incubated at 50 ºC for 5 min. 1 µl (or 200 U) of SuperScript IV Reverse Transcriptase (Thermo Fisher) was added to the +RT reaction tubes and 1 µl of dH2O to the –RT reaction tubes. The mixture was incubated at 50 ºC (or 55 ºC for Section 3.2.2) for 1h, followed by 15 min at 70°C. 1 µl of RNase H (2.5U/µl) was added to the reaction, which was then incubated for 20 min at 37°C, to degrade the RNA in the RNA-cDNA duplexes, which could affect primer annealing and reaction efficiency during the qPCR. Dilutions of 1:4 (in experiments described in Sections 3.1 and 3.3) or 1:8 (in experiments described in 3.2) were made from the RT products. cDNA was stored at -20°C.

2.5.qPCR

2.5.1. Primer Design

Primers were designed to anneal to the subtelomeric DNA sequence of a chromosome end proximal to the telomeric repeats and amplify a 90-150 bp target downstream of the TERRA TSS. Primers were designed with the online software primer3 (https://primer3.ut.ee/). Primer melting temperatures (Tm) were in the 64-69 ºC range. Primers were confirmed not to form cross primer dimers, using Multiple Primer Analyzer (ThermoFisher Scientific)

The target sequences for amplification were aligned with the entire human genome using the basic local alignment search tool (BLAST) at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) to detect possible nonspecific amplification during the qPCR. primer3 is expected to give uniquely binding primers, however, the subtelomeric sequences could potentially be absent from its database. BLASTing against the transcriptome was, therefore, used to gain more confidence in primer specificity.

Five sets of forward and reverse primers were designed for the 7q_12q subtelomeric sequence. Primer efficiency was tested (see Section 2.5.3) and a test qPCR was also run with all forward and reverse primer pair. The 2 primer pairs with highest amplification efficiency closest to 100% were chosen for the 7q_12q final experiments.

Two forward and reverse primer pairs were designed for each of the remaining sequences (5p, 15q, 16p and 20q chromosome ends) and the primer pair with amplification efficiency closest to 100% was chosen for the final experiments. The full list of qPCR primers used in this project can be seen in Table 3. An example of the location of the qPCR target sequences in the subtelomere can be seen in Figure 11, for the 7q_12q consensus sequence.

Table 3: qPCR Primers chosen for use in this study following primer efficiency testing

Primer	Sequence (5’ -> 3´)	Tm (ºC)
5p Fw	CATTTGTCTTCCGAGGCTGC	67,3
5p Rev	CAAAAGCCCCTCTGAATCCTG	67,1
7q_12q Fw 1	ATTTGTCTTCCGAGGCTGC	64,7
7q_12q Fw 2	GAGCAATGGGGTGTGTCATA	63,9
7q_12q Rev	GTTCAGTGTGGAAAACGGGA	65,3
15q Fw	CGATGCTGCAACTGGACCC	69,6
15q Rev	GCGGTTCAGTGCGGAAAAC	68,2
16p Fw	CCCGTTTTCCACACTGAACC	66,9
16p Rev	CTCCTAAGCACAGACGTTGGG	66,2
20q Fw	CAATAATTCGAAAAGCCGGGC	68,3
20q Rev	GGCCACCACTATAAGCAACAG	63,7

Another factor to consider is the possibility of nonspecific amplification. There is a variety of conserved sequences across subtelomeres and designing primers that specifically target only one can be difficult to achieve. Primer specificity of the designed primers was extensively tested, to assure that a primer pairs designed for a specific subtelomere had no full matches in any other. For instance, all qPCR end products were run in a 1% agarose gel and checked for presence of more than one band, denoting nonspecific amplification. Amplicon size was also confirmed.

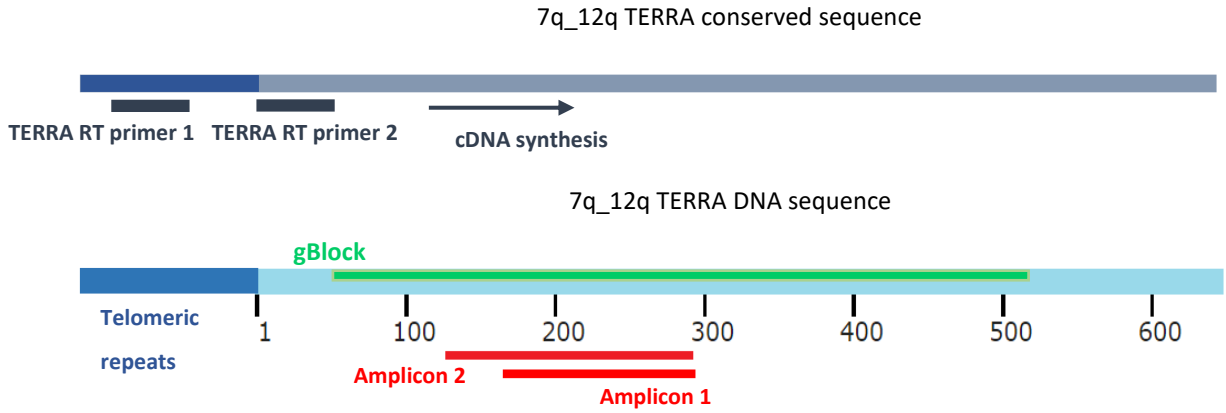

Figure 11: Representation of the reverse transcription reactions of the 7q_12q TERRA sequence aligned with the 7q_12q DNA sequence. Telomeric and subtelomeric derived sequences represented in dark blue or light blue, respectively. The two reverse transcription primers used in the study are represented over their complementary regions in dark blue (bar size does not reflect primer length). Arrows indicate cDNA synthesis direction. Amplicon targets for qPCR and gBlock standard used for quantification are represented in red or green, respectively, over the 7q_12q cDNA. It's possible to observe that cDNA synthesis covers the entirety of the target amplicons and of the region to be used as a standard.

2.5.2. qPCR Reaction

A master mix was prepared for each primer, containing 6 µl dH_2O, 1 µl forward primer (10 µM), 1 µl reverse primer (10 µM), 10 µl 2X Power SYBR Green PCR Master mix (Applied Biosystems). 2ul of these cDNA was added into qPCR reaction tubes, in triplicates and the -RT reactions were added into other tubes. 18 µl of the primer master mix. Tubes were closed and inserted in a Rotor Gene 6000 System (Corbett). qPCR consisted of 5 min at 95°C, followed by 45 cycles at 95°C for 15s and 60°C for 30 seconds

2.5.3. Primer Efficiency Testing

Serial dilutions of gDNA were made over 5 orders of magnitude, in dH_2O, as follows (Table 4):

Table 4: Concentrations of the 5-point serial dilutions of gDNA for primer efficiency testing

No.	Dilution	Concentration (ng/µl)
1	undiluted	10
2	1:100	1
3	1:1000	0.1
4	1:10000	0.01
5	1:100000	0.001

For each primer, qPCR was performed in triplicates on the dilution series of gDNA. Average CTs values for each dilution were calculated. The log value of the starting quantity was calculated. The slope of the regression curve between the log values and the average CT values was calculated in Microsoft Excel. Primer Efficiency was calculated using formula 1:

$$1. \quad E = \left(10^{\frac{-1}{slope}} - 1\right) \times 100$$

The primer pairs with efficiency values closer to 100% was selected for each chromosome to use in the final experiment (as previously stated, 2 primer pairs were selected for 7q_12q).

2.6.TERRA Absolute Quantification

2.6.1. Standard Curve – Calculating TERRA Copy number

To measure the number of TERRA-derived cDNA molecules in the cDNA sample, absolute quantification was performed. For this, a standard curve was made with a standard in known concentrations. Standards were designed to be identical to 5 distinct 300-500 bp subtelomeric regions (5p, 7q_12q, 15q, 16p, 20q), as close as possible to the 5'-(CCCTAA)-3' repeats and downstream of the TERRA TSS predicted by ONT-seq.

Thus, each 300-500 bp standard contains the 90-150 bp fragment targeted for amplification by qPCR and is identical to the cDNA obtained from TERRA transcribed in that chromosome end. Standards were ordered as gBlocks Gene Fragments from Integrated DNA Technologies. The gBlock sequences were cloned into a pJET1.2/Blunt plasmid (Thermo Scientific pJET Cloning Kit) and verified by Sanger Sequencing (Lighrun Tube sent to Eurofin Genomics). The plasmid concentrations were measured by Nanodrop. The list of gBlocks used as standards for absolute quantification in this study can be seen in supplemental Table A 1.

The plasmids containing the gBlocks were used to create a 5-point standard curve. The number of standard copies in each dilution was calculated using formula 2:

$$2.\quad Standard\ copy\ number = \frac{6.02\times1023(\text{copies mol}^{-1})\times \text{DNA amount(g)}}{\text{DNA length (bp)}\times 660\ (\text{g mol}^{-1}\text{bp}^{-1}\ )}$$

The standard curve dilutions were as follows (Table 5):

Table 5: Copy numbers of the 5-point serial dilutions of standard used for absolute quantification

Dilution No.	Standard copies per reaction
1	10^6
2	10^5
3	10^4
4	1000
5	100

The standard dilution series was amplified by qPCR at the same time as TERRA cDNA. The average CT values of the standard dilutions were determined. A regression curve was made between the log values of the copy numbers in the standard dilutions and the average CT values standard dilutions was calculated in Microsoft Excel. Finally, the CTs of the target samples were plotted on to this standard curve to obtain TERRA copy numbers

2.6.2. Calculating Copy Number per Thousand Cells

To calculate the copy number per cell, cells were counted before RNA extraction and the total RNA yielded from the extraction was quantified with Nanodrop. Thus, the ratio of RNA yielded per thousand cells, for each cell line, was calculated. Known RNA mass (4 to 5 ug) was used in the RT reaction and RT reaction was assumed to be 100% efficient, allowing for estimation of the number of cells corresponding to the nanograms of cDNA present in one qPCR tube. Finally, copy number per thousand cells was calculated with formula 3:

$$3.\quad \text{Copy number per thousand cells} = \frac{TERRA\ Molecules\ per\ qPCR\ reaction}{Thousands\ of\ cells\ corresponding\ to\ nanograms\ of\ cDNA\ in\ qPCR\ reaction}$$

2.7. DNA Sequence Analysis

To analyze CpG dinucleotide contents and predict CpG islands, the CpGPlot/CpGReport at the European molecular biology open software suite program (EMBOSS) (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/) [94] was utilized. The results were validated using the CpG function of the MethPrimer 2.0 tool at Li Lab (Peking Union Medical College Hospital) (http://www.urogene.org/cgi-bin/methprimer2/MethPrimer.cgi) [95]

2.8. DNA Methylation Analysis

Genomic DNA was extracted from HCT116 and HCT116 DKO cells. 5 ug of DNA from both cell lines was digested in 40 µl total volume, with 1 µl (or 20 U) of HpaII (CpG methylation-sensitive enzyme), in 1X Cutsmart Buffer, overnight. Another 5 ug of DNA from both cell lines was digested in identical conditions with MspI (CpG methylation-insensitive enzyme). The digested DNA was run on a 0,8% agarose gel and an image was taken with ChemiDoc XRS+ System (Bio-Rad).

2.9. 7q and 12q Subtelomeric Sequence Isolation by PCR

Genomic DNA was extracted from U2OS, HeLa, HEK and SAOS2. Using primer3, a single forward primer and two distinct reverse primers were designed to amplify the 7q subtelomeric CpG island sequence and the 12q subtelomeric sequence (Table 6). Using these two primer pairs sharing the same forward primer, two partially overlapping subtelomeric sequences were targeted for amplification, one being approximately 200 bp longer than the other. The two sets of primer pairs did not allow to distinguish between 7q and 12q sequences, due to the high homology between the sequences. Instead, 7q and 12q sequences were distinguished from each other after amplification by the presence of a 30 bp sequence in the 12q sequence that is not present in the 7q sequence.

The following reaction was set up: 10 µl of HF or GC Buffer (5X), 0,4 µl of dNTP mix (25 mM each), 0,5 µl of Forward Primer (100 mM), 0,5 µl of Reverse Primer (100 mM), 0,5 µl of Phusion Polymerase Enzyme (2 U/ml), 50 ng of template DNA and filled to 50 µl with dd H_2O

Two sets of reaction were set up for target sequence, one using HF Buffer and one with GC Buffer. Three different temperatures for primer annealing during the qPCR were also tested, 68 ºC, 70 ºC and 72 ºC. PCR reactions consisted of 98 °C for 30 s (1 cycle), 98 °C for 10/15 s (30-45 cycles), 72 ºC for 20 sec (30-45 cycles), 68/70/72 °C for 54 s (30-45 cycles) and 72 °C for 5 min (1 cycle)

The amplified sequences were cloned into a pJET1.2/Blunt plasmid (Thermo Scientific pJET Cloning Kit) and verified by Sanger Sequencing (Lightrun Tube sent to Eurofin Genomics). The plasmid concentrations were measured by Nanodrop.

Table 6 oligos used for PCR amplification of 7q and 12q CpG islands

Oligo	Sequence (5' -> 3´)
7q_12q_CpG_island_FW1	ATGAGCAATGGGGTGTGTTATATTTTCGGTGTCAT
7q_12q _CpG_island_RV1	GGGTGGCATATTTTGGTCTTATACACTGTGTTCCA
7q_12q _CpG_island_RV2	ATGTCGACAGAGGTAGCTTTTAAATGGGGATTTCC

2.10. Plasmid Constructions

Cloning of gBlocks and of PCR amplifying sequences into pJET1.2/Blunt plasmid (Thermo Scientific pJET Cloning Kit) was performed as follows: Ligation reaction was set up on ice, containing 10 µl of 2X Reaction Buffer, 0.15 pmol ends of insert (calculated using Promega online Biomath Calculator), 1 µl (0.05 pmol ends) of pJET1.2/Blunt plasmid, 1 µl T4 DNA Ligase and filled to 20 µl with dH_2O. The ligation mix was incubated at room temperature for 15 min. 5 µl of the ligation mix were added to 30 µl Stbl3 *E. Coli* in a 1.5 ml Eppendorf and chilled on ice for 20 min. Thermal shock was applied at 42 ºC for 45 s. Bacteria were left on ice for 2 more min, then 200 µl of LB was added. Bacteria recovered in LB at 37 ºC with 600 rpm rotation for 35 min. 100 µl were added to a LB agar plate with ampicillin resistance. Colonies were left to grow overnight at 37 ºC.

Two colonies were picked and grown in liquid LB + ampicillin culture in an incubator agitating at 220 rpm. Lastly, plasmids were extracted from bacteria using NYZ MiniPrep. Concentrations were measured using Nanodrop. Insert sequences were confirmed by Sanger Sequencing (Lightrun Tube sent to Eurofin Genomics)

2.11. CRISPRi Assay

2.11.1. sgRNA Design

20 bp sgRNAs were designed with CRISPR-ERA webtool (http://crispr-era.stanford.edu/). For each of the 7q and 12q sequences, a pairs of primers was designed that annealed with each other, forming a duplex, while leaving 4 basepair long single stranded ends on each side. These ends are complementary to BsaI cuts that result from digestion of the CRISPRi empty vector. sgRNAs were annealed to each other by adding, in a 1.5 ml Eppendorf, 1 µl forward sgRNA (0,5 µg/µl), 1 µl reverse gRNA and 48 µl TE (pH=8.0). This reaction mixture was incubated at 95 ºC for 5 min and then cooled slowly to room temperature. The sgRNAs designed for insertion into the CRISPRi plasmid can be seen in supplemental Table A 2

2.11.2. Plasmid Construction

The backbone plasmid to be used for CRISPRi depletion strategy was gifted to Azzalin lab by Prof. Ulrike Kutay. It already contained the deactivated Cas9 protein (dCas9), KRAB repressor and sgRNA scaffold. 2 µg of backbone plasmid were digested with BsaI-HF enzyme in a 20 µl total volume. The linearized plasmid was purified using Zymoclean Gel DNA Recovery Kit. Plasmid was eluted in 20 µl dd H2O.

Duplex sgRNAs were cloned into the backbone dCas9 plasmid between BsaI sites, in Stbl3 E. Coli. For this, the following reaction mixture was added in a 1.5 ml Eppendorf: 50 µg of digested backbone plasmid, 2 µl of ds sgRNA mix, 2 µl T4 DNase Ligase Buffer, 1 µl T4 DNA Ligase, to 20 µl with dd H2O

The ligation mix was left at room temperature for 2 hours. 10 µl of the ligation mix was added to 50 µl of Stbl3 *E. Coli*. The rest of the cloning was performed as described in Section 2.10, except bacteria recovered after thermic shock in 500 µl total volume instead of 235 µl.

2.11.3. Cell Transfection

Plasmids were transfected into U2OS cells using Lipofectamine RNAiMAX (Invitrogen). 6 µg of plasmid and 15 µl of Lipofectamine were separately diluted in Opti-MEM (Thermo Fisher Scientific). The two solutions were added together, briefly mixed by pipetting and incubated at room temperature for 10 min.

These mixtures were then added to 10cm plate cell cultures. Separate transfections were performed using each of the plasmids in Table 7.

Table 7: Plasmids constructed for the CRISPRi TERRA depletion assay

Name	Description
CRISPRi_7q	dCas9 vector with 7q sgRNA inserted
CRISPRi_12q	dCas9 vector with 12q sgRNA inserted
CRISPRi_EV	dCas9 empty backbone vector

Medium was changed 5 h after adding the transfection mixture. Puromycin was added to a final concentration of 1 µg/mL, 24 h after selection and to a “mock” U2OS culture that had not undergone transfection. 48 h after selection, there was complete cell death in the “mock” culture. At this point total RNA was isolated from 1,3 to 1,9 million cells transfected with each of the plasmids, as described in 2.2. RT-qPCR was performed as described in 2.4, 2.5.

2.12. Statistical Analysis

Two-tailed Student's t-test was performed in Microsoft Excel for a direct comparison of two groups. One-way analysis of variance (one-way ANOVA) was employed in GraphPad Prism 8 for comparison of more than two groups, followed by Tukey's HSD treatment to adjust the pairwise statistical significance. *P* values are indicated as: $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$. Error bars represent SD.

3. Results

3.1. Validation of 7q_12q TERRA Quantification as Diagnostic Tool

In the first step of this project, the goal was to test quantification of TERRA from the long arms of chromosomes 7 and 12 as a possible diagnosis tool for ALT cancers. The working hypothesis was that, for 7q_12q TERRA quantification as diagnosis to be viable, 7q_12q TERRA levels in ALT cells should be, at the minimum, 50 to 100-fold higher than in telomerase-positive cell lines. An increase of this magnitude would allow for the establishment of a diagnosis threshold that ensured reliability and sensitivity (further discussed in Section 4.1).

TERRA from chromosomes 7q and 12q (7q_12q TERRA) was quantified by RT-qPCR and compared between a human ALT osteosarcoma cell line (U2OS) and a human telomerase-positive osteosarcoma cell line (HOS). Absolute and relative quantification of TERRA was performed in triplicate (Figure 12). Because of the extremely high similarity between the 7q and 12q subtelomeric sequences, the qPCR primers were designed to amplify the conserved sequence between the two chromosome ends (7q_12q sequence) and did not allow distinguishing between the two ends.

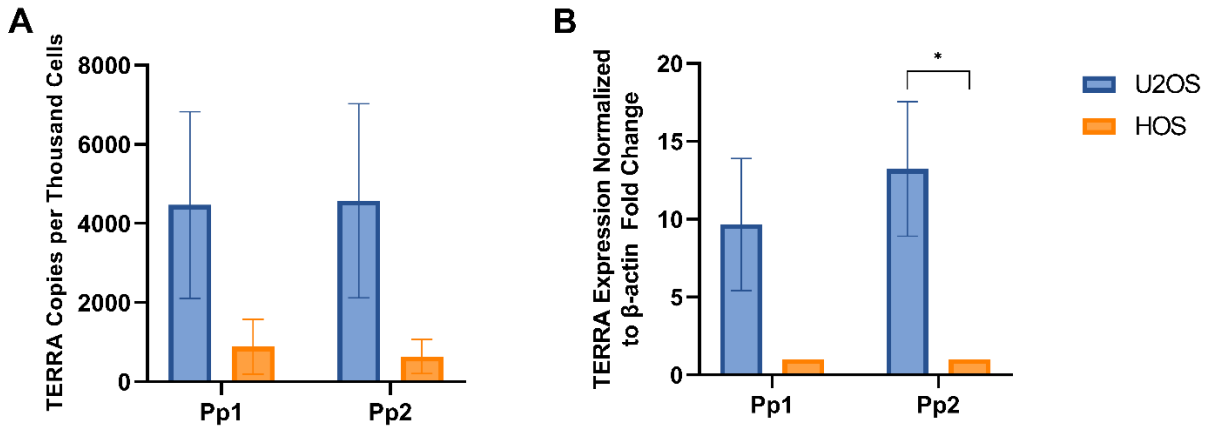

*Figure 12: qPCR quantification of TERRA from chromosomes 7q and 12q inU2OS and HOS cells. Reverse transcription performed with primer o1 (C-rich RT primer from Table 2). (A) Absolute quantification. TERRA Copies per Thousand Cells displayed on y axis. (B) Relative quantification. All results normalized to β-actin. Fold Change in TERRA expression normalized to β-actin displayed on y axis. Bars and error bars represent means and SDs from three independent experiments. P values were calculated with a two-tailed student's t-test. *P ≤ 0.05*

Considering the absolute quantification results (Figure 12A), U2OS cells expressed an average of about 4500 7q_12q TERRA copies per thousand cells and telomerase-positive HOS cells expressed about 900. The results were similar for both sets of qPCR primers. While 7q_12q TERRA expression is clearly higher in U2OS than HOS cells, the fold increase in U2OS relative to HOS is less than ten-fold. The variability between biological replicates is also high in both U2OS and HOS.

As for the relative quantification (Figure 12B) - which was performed by normalizing TERRA to the β-actin housekeeping gene - a 10-fold increase in 7q_12q TERRA is observed in U2OS relative to HOS, using primer pair 1 and about 13-fold increase using primer pair 2. The variability between biological replicates in both U2OS and HOS remains quite elevated.

A possible source of bias in these results comes from the choice of primer for cDNA synthesis during the reverse transcription step. The reverse primer was designed to be complementary to the TERRA (5'-UUAGGG-3') sequence derived from the telomeric 5'-(CCCTAA)-3' repeats (C-rich RT primer o1 from Table 2, Section 2.4.1), which were assumed to be identical in both U2OS and HOS cell lines. However, it has been reported that ALT cells often have degenerated repeats at their telomeres [96]–[98]. Thus, reverse transcription from U2OS TERRA could be less efficient when compared to TERRA from HOS. Such difference in reverse transcription efficiency could bias the final qPCR results, leading to an underestimation of U2OS TERRA expression.

To test the existence of this bias, the RT-qPCR was repeated in identical conditions but using a reverse transcription primer that annealed to the subtelomeric part of the 7q_12q TERRA transcripts (7q_12q subtelomeric RT primer o2 from Table 2, Section 2.4.1), right upstream of the telomeric repeats, assuring that cDNA still covers the entire qPCR target region for amplification. The results of the second set of RT-qPCRs are shown in Figure 13.

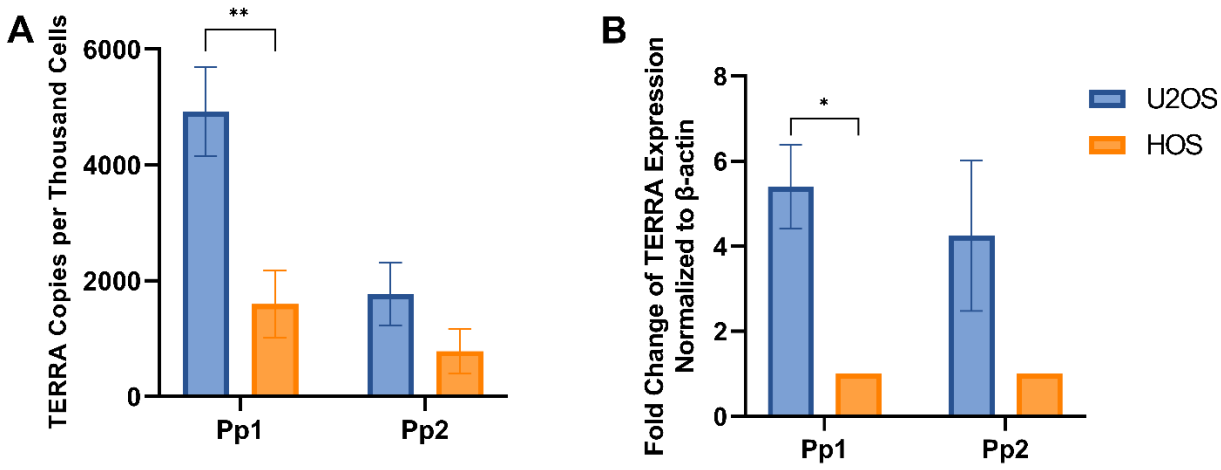

*Figure 13: qPCR quantification of TERRA from chromosomes 7q and 12q inU2OS and HOS cells. Reverse transcription performed with primer o2 (7q_12q subtelomeric RT primer from Table 2). (A) Absolute quantification. TERRA Copies per Thousand Cells displayed on y axis. (B) Relative quantification. All results normalized to β-actin. Fold Change in TERRA expression normalized to β-actin displayed on y axis. Bars and error bars represent means and SDs from three independent experiments. P values were calculated with a two-tailed student's t-test. *P ≤ 0.05, **P ≤ 0.01*

Considering the absolute quantification (Figure 13A), HOS contain about 1600 7q_12q TERRA copies per thousand cells with primer pair 1, and 800 copies with primer pair 2. The results are similar to those using the C-rich primer o1n Figure 12A (600 copies per thousand cells), indicating that the two primers prime reverse transcription with similar efficiencies for HOS 7q_12q TERRA.

7q_12q U2OS TERRA expression was also similar for primer pair 1 (about 5000 copies per thousand cells), when compared to the results using the C-rich RT primer (4500 copies per thousand cells). For primer pair 2, a small decrease occurred, with TERRA levels being around 1800 copies per thousand cells.

The relative quantification results (Figure 13B), also displayed a lower 7q_12q TERRA expression fold using the 7q_12q subtelomeric RT primer, when compared to the C-rich RT primer. A 5-fold 7q_12q TERRA increase was observed for primer pair 1 and 4-fold increase for primer pair 2.

Overall, very similar 7q_12q TERRA levels were observed using cDNA reverse transcribed using the C-rich telomeric RT primer compared to the 7q_12q subtelomeric derived reverse prime. This suggests that reverse transcription efficiency is not affected by primer choice. Therefore, either there is no significant degeneration of U2OS telomeric repeats, or, if there is, the RT can still efficiently reverse transcribe TERRA from these telomeres.

The highest 7q_12q TERRA fold-increase in U2OS when compared to HOS was of 13-fold. This is much lower than the 50-100-fold minimum that has been hypothesized to be necessary for establishment of a diagnosis threshold. This, combined with the high variability between biological replicates in both cell lines, indicates that establishing a diagnosis threshold for quantified 7q_12q TERRA expression for a sensitive clinical tool may be unfeasible. This matter is further discussed in Section 4.1.

3.2. Validation of ONT-seq Results by RT-qPCR

The second goal of this project was a broader validation of the ONT-seq results in a panel of 4 cell lines: U2OS and SAOS2 ALT cells, and HeLa and HEK293T telomerase-positive cells. Five chromosome ends were chosen for validation, 5p, 7q_12q,15q,16p and 20q. These 5 ends represent a range of TERRA expression values, from high (7q and 12q) to moderate (5p and 16p) to low (15q and 20q). The ONT-seq results for these chromosome ends are shown in Figure 14.

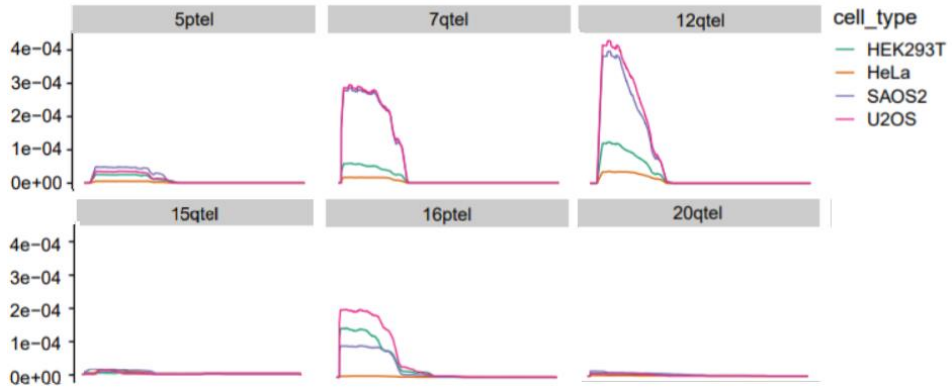

Figure 14: Oxford ONT-seq of TERRA expression for 5p, 7q, 12q,15q, 16p and 20q chromosome ends, for U2OS, HeLa, SAOS-2 and HEK293T. x axis: first 2kbp of subtelomere after telomeric repeats. y axis: TERRA read coverage in log scale

TERRA levels were measured by RT-qPCR by absolute quantification. The RT-qPCR validation results are shown in Figure 15.

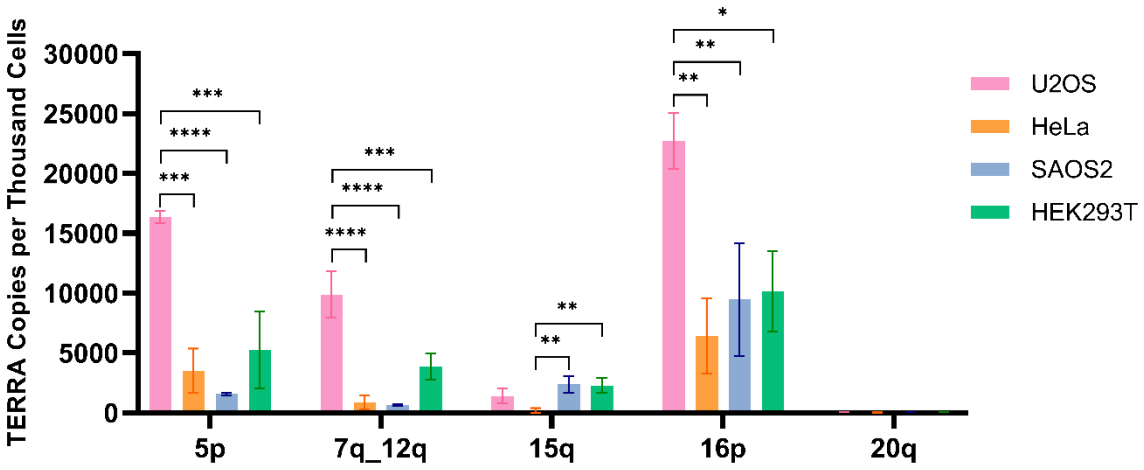

*Figure 15: qPCR quantification of TERRA from chromosomes 5p, 7q, 12q, 15q, 16p and 20q in U2OS, HeLa, SAOS-2 and HEK293T cells. Absolute quantification. TERRA Copies per Thousand Cells displayed on y axis. Bars and error bars represent means and SDs from three independent experiments. P values were calculated with a one-way ANOVA test. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001*

Considering the RT-qPCR (Figure 15), chromosome ends 5p and 16p showed higher TERRA expression across all cell lines, when compared to the ONT-seq results, even surpassing 7q_12q TERRA levels. However, the standard curves obtained from these two chromosome ends were less linear, with a low R^2 value (0.93 for 5p and 0.9 for 16p). These curves were also not representative of the 1:10 dilutions made. A 10-fold difference in concentration should result in a 3.3 CT value difference between curves. However, the points in the 5p and 16p curves were, at

times, separated by 5 CTs. Therefore, the results for these two chromosome ends are less reliable. and likely biased these results. The standard curves for each of the chromosome ends can be seen in supplemental Figure A 1. Reasons for lack of linearity and representative value in standard curves are discussed in Section 4.2.

As for the remaining chromosome ends, 7q_12q, 15q and 20q, the results are similar to the ONT-seq with regard to U2OS, HEK2937 and HeLa. However, the outlier was SAOS2 osteosarcoma cells, which, despite being an ALT cell line, had much lower TERRA expression than predicted by ONT-seq, particularly in 7q_12q ends. SAOS2 TERRA is also underexpressed in 5p and 16p compared to U2OS. Northern blot analyses of total TERRA performed in the Azzalin lab have also showed that SAOS2 have considerably higher overall levels of TERRA than HeLa and HEK293T, although lower than U2OS. Considering the possibility that low TERRA levels in SAOS2 were a result of bias caused by different primer efficiencies or reverse transcription efficiencies between cell lines, a series of controls were performed to test these causes for bias.

3.2.1. qPCR Primer Validation

A possible reason for low detection of TERRA expression in SAOS2 is inefficient annealing of the qPCR primers to the SAOS2 subtelomeric sequence in comparison to the other cell lines. This could be caused by nucleotide polymorphisms in the SAOS2 target sequences. Hence, primer annealing to the DNA sequence was validated by performing a qPCR with genomic DNA from each of the 4 cell lines. QPCR was performed with equal amounts of genomic DNA from each cell line. For each telomere end, if primer annealing efficiencies are similar across cell lines, DNA amplification for all cell lines should be similar as well.

Primer annealing validation results can be seen in Figure 16. Relative quantification was performed, with target DNA amplification being normalized to β-actin. Four of the previously used primers were tested.

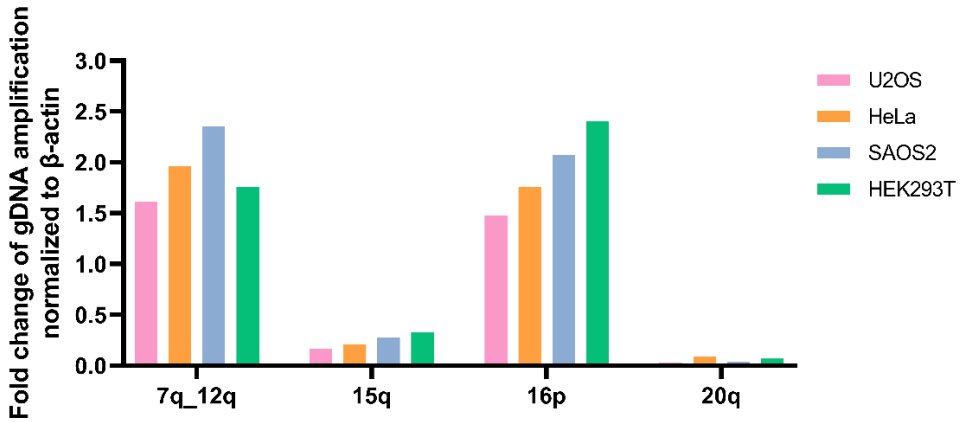

Figure 16: qPCR primer validation in U2OS, HeLa, SAOS2 and HEK293T for 5p, 7q_12q,15q,16p and 20q chromosome ends. Relative quantification. All results normalized to β-actin. y axis: Fold Change in gDNA amplification, normalized to β-actin.

For each telomere end, there is little variation in the fold change across cell lines, including SAOS2. This indicates that there is no significant bias caused by inefficient primer annealing to the SAOS2 subtelomeric sequence.

3.2.2. Reverse Transcription Validation

A second possible source of bias, as previously mentioned in Section 3.1, is a lower SAOS2 TERRA reverse transcription efficiency compared to other cell lines. This could be caused by poor RT primer annealing to degenerated SAOS2 telomeric repeats, leading to underestimation of TERRA expression. This was tested, as in Section 3.1, by repeating the RT-qPCR for one sample RNA of each cell line using a primer designed for the TERRA subtelomeric sequence (primer o2 from Table 2). The results of this control (Figure 17A), represented with TERRA levels in all cell lines normalized to HeLa TERRA levels, show a very similar outcome to the RT-qPCR, with SAOS2 expressing TERRA in the same range as telomerase-positive HeLa and HEK293T cells, and far below U2OS, meaning that low SAOS2 TERRA expression cannot be justified by a bias caused by RT.

A third possible reason for SAOS2 TERRA levels concerned the RT reaction temperature. The RT to make cDNA for the ONT-seq was performed at 55 ºC, but the RT used for the qPCR validation was performed at 50 ºC. In the case that the SAOS2 subtelomeric sequence contains more secondary structures than the remaining cell lines, a bias may have been caused by the ability of the RT to unwind these secondary structures at 55 ºC but not at 50 ºC. This was tested by repeating the RT-qPCR for one sample RNA of each cell line and performing the RT at 55 ºC only for 7q_12q primers. The results of this experiment can be seen in Figure 17B, and they are highly similar to the results with RT at 50 ºC. Thus, this possible reason for low SAOS2 TERRA expression was ruled out.

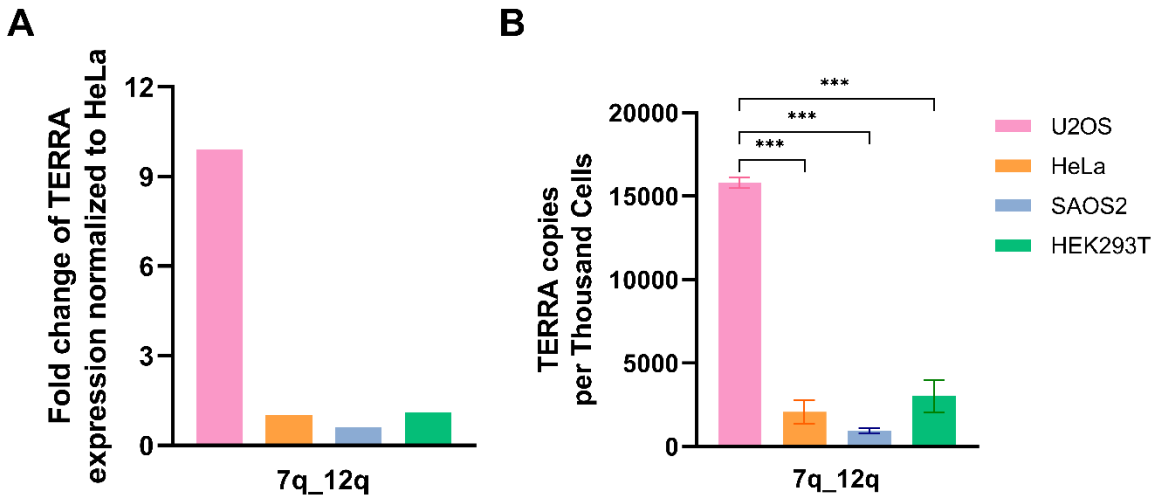

*Figure 17: qPCR quantification of TERRA from chromosomes 7q and 12q in U2OS, HeLa, SAOS2 and HEK293T cells. (A) Reverse transcription performed with primer o2 (7q_12q subtelomeric RT primer from Table 2). Relative quantification. All results normalized to ß-actin. Fold Change in TERRA expression normalized to ß-actin displayed on y axis. (B) cDNA from reverse transcription at 55 ºC. Absolute quantification. TERRA copies per Thousand Cells represented on the y axis Absolute quantification. TERRA Copies per Thousand Cells displayed on y axis. Bars and error bars represent means and SDs from three independent experiments. P values were calculated with a two-tailed student's t-test. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001*

More possibilities for the discrepancy between the ONT-seq and RT-qPCR SAOS2 results will be discussed in the Section 4.2. Overall, the ONT-seq results were partially validated. One element that is confirmed by the RT-qPCR is the high 7q_12q TERRA expression in ALT U2OS cells, despite these chromosome ends not containing the previously characterized 61-29-37 repeat promoter, which leads to the next section of this project: the analysis of the 7q_12q conserved subtelomeric sequence, its promoter activity, and transcriptional regulators.

3.3.7q_12q TERRA Transcription Regulation by DNA Methylation

The first step taken to study active TERRA transcription by 7q and 12q subtelomeric CpG island putative promoter sequences was to study the impact of DNA methylation on their promoter activity. ALT cell subtelomeric CpG islands have been shown to be hypomethylated [16], and increased transcription of TERRA in ALT cells is believed to be a result of this TERRA promoter hypomethylation [66]. Nergadze et al. analyzed the methylation state of 61-29-37 repeats in human HCT116 cells knocked-out for DNMT1 and DNMT3b (double KO [DKO]) DNA methyltransferases [16]. Concomitant disruption of DNMT1 and DNMT3b genes was shown to completely abolished methylation at 61-29-37 repeat CpG dinucleotides, implying that DNMT1 and DNMT3b enzymes cooperatively maintain DNA methylation at 61-29-37 repeats in HCT116 cells. This decreased methylation state resulted in TERRA upregulation in 61-29-37 bp containing subtelomeres in HCT116 DKO cells [16]. In this way, DNA methylation has been shown to be a crucial suppressor of TERRA expression from the previously characterized 61-29-37 repeats TERRA promoters.

For this study, the hypothesis was that a decreased methylation state could have the same impact on the 7q_12q putative promoter sequence. As such, HCT116 cells and HCT116 DKO cells were chosen to study this effect in the 7q and 12q chromosome ends. RT-qPCR was performed with 7q_12q primers on the parental telomerase-positive cancer cell line HCT116 and HCT116 DKO cells. Absolute (Figure 18A) and relative (Figure 18B) quantification were performed.

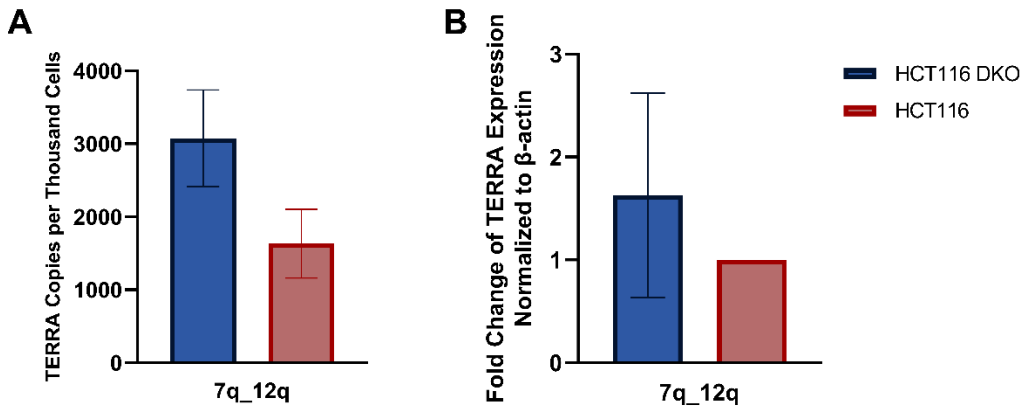

Figure 18: qPCR quantification of TERRA from chromosomes 7q and 12q in HCT116 and HCT116 DKO cells. Reverse transcription performed with primer o1 (C-rich RT primer from Table 2). (A) Absolute quantification. TERRA Copies per Thousand Cells displayed on y axis. (B) Relative quantification. All results normalized to β-actin. Fold Change in TERRA expression normalized to β-actin displayed on y axis. Bars and error bars represent means and SDs from three independent experiments. P values were calculated with a two-tailed student's t-test.

Both the absolute (Figure 18A) and relative (Figure 18B) quantification show that DKO hypomethylated cells express more TERRA from 7q and 12q telomeres relative to the parental cells. However, only a very minor increase, about 2-fold in absolute quantification and 1.5-fold in relative quantification, was observed when

compared to previous studies with these cells lines in other chromosome ends [16], [18]. For instance, Lingner et al. reported a 20-fold increase in DKO TERRA expression relative to HCT116 at 12q chromosome end normalized to GAPDH.

To validate the cell lines, an experiment was performed to verify the high methylation state of HCT116 cells and demethylated state of DKO cells (Figure 19B). Genomic DNA from both cell lines was digested with the methylation-sensitive HpaII restriction enzyme or with its methylation-insensitive isoschizomer MspI, and the digested DNA was run in an agarose gel. HpaII digested HCT116 DNA displays a prominent genomic DNA band at high molecular weights, while HpaII digested DKO DNA displays only a smear. The MspI digested controls both display a smear since MspI digests both methylated and non-methylated DNA. This confirms that HCT116 cell DNA is methylated in vivo, but not DKO cell DNA.

The 7q_12q result was controlled by performing RT-qPCR for TERRA expression on a chromosome end known to have a CpG island with TERRA 61-29-37 bp repeats promoter, positive control XqYq, and one known to not contain a CpG island, negative control XpYp (Figure 19A) (primers used for RT-qPCR in these ends can be seen in supplemental Table A 3). However, TERRA fold-increase between HCT116 and DKO cells, for both chromosome ends, was very minor.

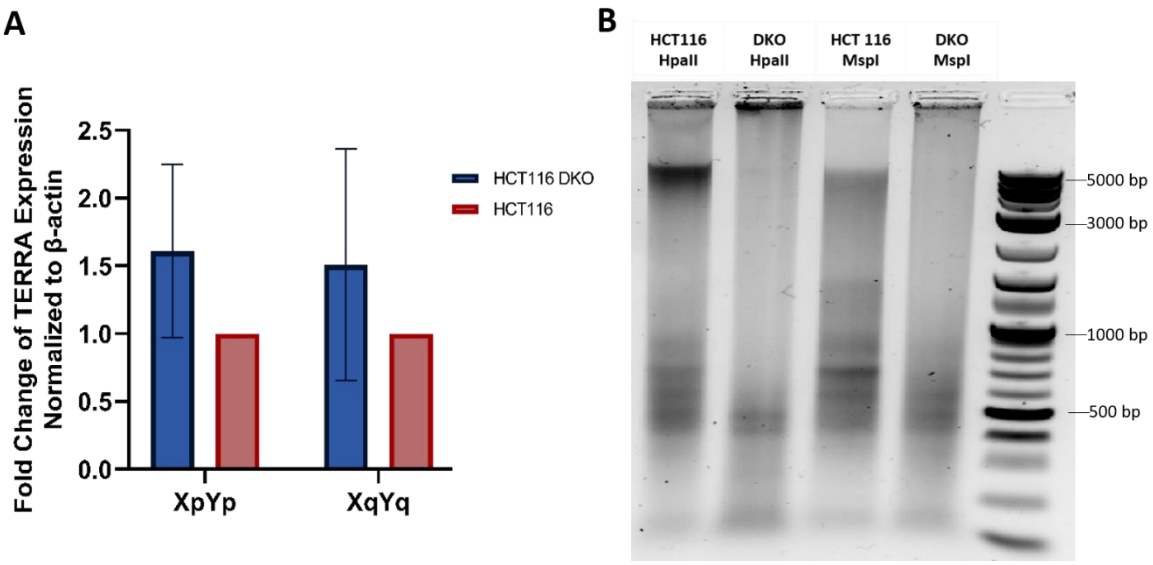

Figure 19: (A) qPCR quantification of TERRA from chromosomes XpYp and XqYq in HCT116 and HCT116 DKO cells. Reverse transcription performed with primer o1 (C-rich RT primer from Table 2). Relative quantification. All results normalized to β-actin. Fold Change in TERRA expression normalized to β-actin displayed on y axis. Bars and error bars represent means and SDs from three independent experiments. P values were calculated with a two-tailed student's t-test. (B) 0,8% agarose gel run of HCT116 and HCT116 DKO gDNA digested with methylation sensitive HpaII enzyme or methylation insensitive isoschizomer MspI

The XpYp and XqYq controls suggest that there is an unforeseen irregularity in these HCT116 and DKO populations and definitive conclusions cannot be taken from the RT-qPCR result. This assay should be repeated with new populations of a wild-type cell line and cell lines knocked out for DNA methyltransferases and DNA methylation regulation of TERRA expression in 7q and 12q chromosome ends still requires further investigating.

3.4.7q and 12q Putative Promoter Isolation

The next objective of the 7q and 12q putative promoter study was to isolate the 7q and 12q subtelomeric sequences for functional assays to be performed. Bioinformatic analysis was performed using the CpG search function of the MethPrimer 2.0 tool (Figure 20A and Figure 21A) and the CpGPlot/CpGReport at the European molecular biology open software suite program (Figure 20B and Figure 21B). The analysis predicted the existence of a 332 bp CpG island in the 7q subtelomere 340 bp away from the first telomeric repeats, and a 249 bp CpG island in 12q located 362 bp away from the first telomeric repeats.

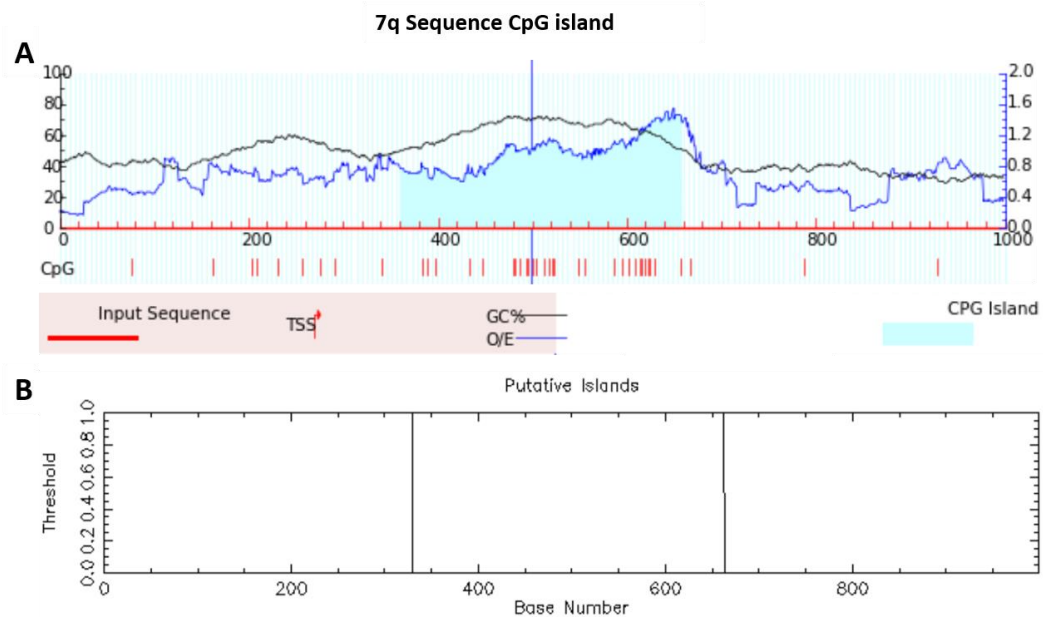

Figure 20: First 1000 bp of the 7q subtelomere immediately following the telomeric repeats analyzed using two bioinformatics tools (A) CpG function of the MethPrimer 2.0 tool. Putative CpG island marked in blue. Predicted TSSs indicated as red vertical bars. CpG island defined as sequence over 100bp long with >50% GC content and >0.6 Observed/Expected Ratio (B) CpGPlot/CpGReport at the European molecular biology open software suite program. Putative CpG island indicated between too vertical black bars. CpG island defined as sequence over 200bp long with >50% GC content and >0.6 Observed/Expected Ratio

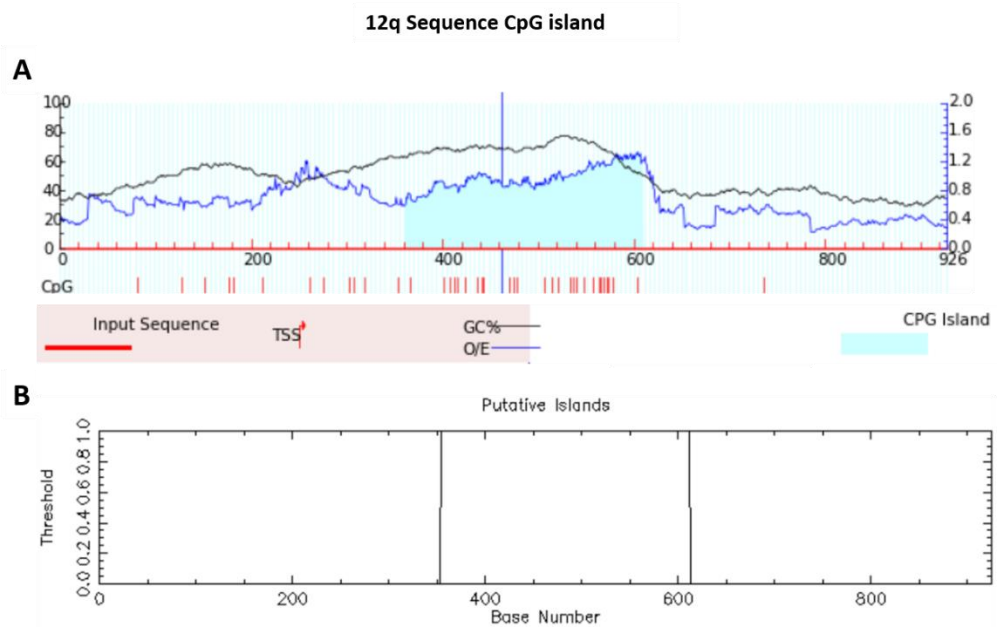

Figure 21: First 1000 bp of the 12q subtelomere immediately following the telomeric repeats analyzed using two bioinformatics tools (A) CpG function of the MethPrimer 2.0 tool. Putative CpG island marked in blue. Predicted TSSs indicated as red vertical bars. CpG island defined as sequence over 100bp long with >50% GC content and >0.6 Observed/Expected Ratio (B) CpGPlot/CpGReport at the European molecular biology open software suite program. Putative CpG island indicated between too vertical black bars. CpG island defined as sequence over 200bp long with >50% GC content and >0.6 Observed/Expected Ratio

The positions of the 7q and 12q CpG islands is represented in Figure 22. The TERRA TSSs as predicted by the ONT-seq experiment are located within the CpG islands, as expected.

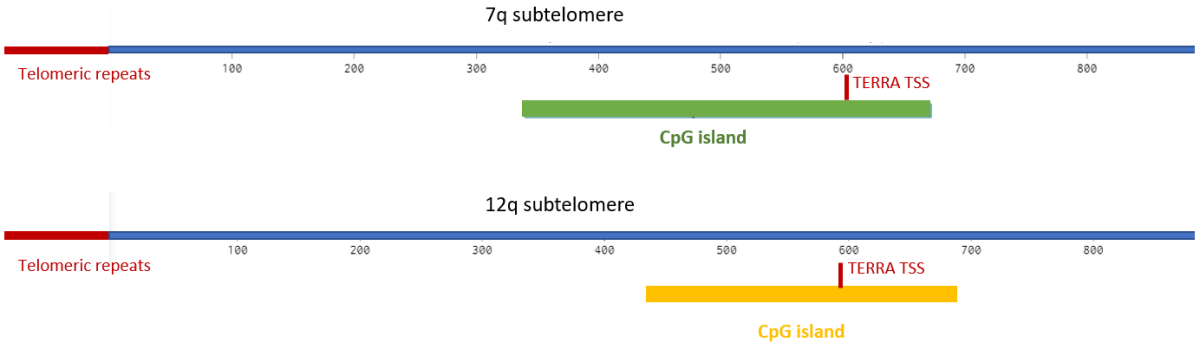

Figure 22: Representation of a portion of the 7q and 12q subtelomeric region. Subtelomeric region indicated in blue and telomeric region indicated in red.7q CpG island indicated in green, and 12q CpG island indicated in yellow. TERRA TSSs predicted by ONT-seq indicated as red vertical bars.

To isolate each sequence, a 697 bp 7q amplicon containing the 332 bp CpG island and a 912 bp 12q amplicon containing the 249bp CpG island were amplified by PCR using genomic DNA from U2OS and HeLa cells (Table 8) (Figure 23). An array of combinations of PCR buffers and primer annealing temperatures were tested in order to find the ideal parameters for amplification. Phusion GC Buffer, which is optimized to improve performance of Phusion DNA Polymerase on GC-rich templates, or templates with complex secondary structures, proved to be

the most efficient buffer. Furthermore, 70 ºC was the only annealing temperature that resulted in amplification, likely because it is high enough to resolve secondary structures yet below the primer annealing temperature.

7q and 12q CpG island-containing sequences were amplified from each cell line. A 7q amplicon from HeLa and a 12q amplicon from U2OS were chosen for cloning into the pjet1.2 plasmid. Sequences were confirmed by Sanger sequencing (Eurofin Genomics).

Table 8: 7q and 12q CpG islands size and locations and size of PCR amplicons containing the CpG islands

Chromosome End	CpG Island Location (bp from telomeric repeats)	CpG Island Size (bp)	Amplicon Size (bp)
7q	340 – 672	332	697
12q	362 – 610	249	912

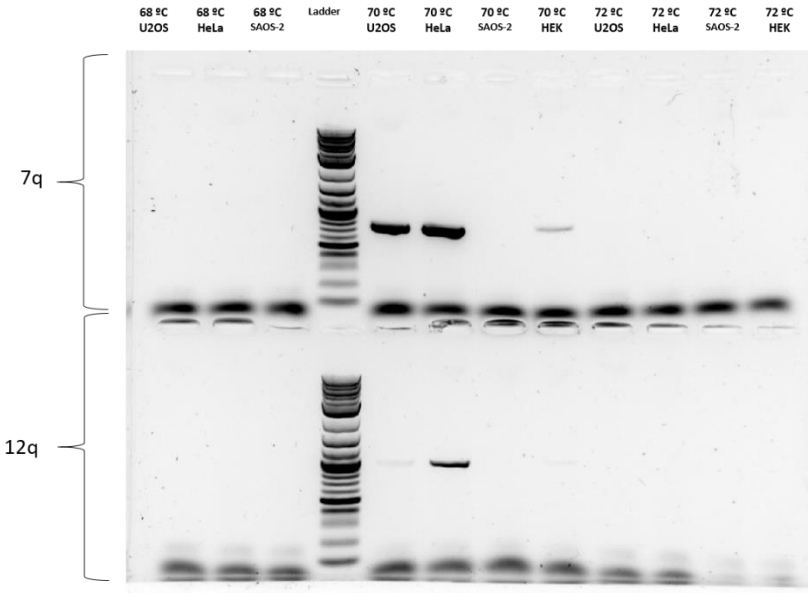

Figure 23: 1% Agarose gel electrophoresis of PCR 7q and 12q amplicons from U2OS, HeLA and HEK cells

Having isolated the sequences containing the 7q and 12q CpG islands and putative promoters, there is a currently ongoing assay being performed in the Azzalin lab to clone both sequences into a reporter plasmid upstream of an eGFP cDNA. This will allow to test for sequence promoter activity. The new data gained on TERRA CpG islands in 7q and 12q can also serve as a foundation for a variety of future assays, such as a knockout of the CpG sequence in 7q and/or 12q and subsequent analysis of telomere length and DNA damage response. Determining the locations of the TERRA TSSs by ONT-seq is also of great use. Bioinformatic analysis have also been performed by collaborators of the Azzalin lab on the sequences upstream of the TERRA TSSs, to locate transcription factor binding motifs. The results of these analysis, and future work, will be further discussed in Section 4.4

3.5.7q_12q TERRA Depletion by CRISPRi Assay

The last attempt to validate the 7q and 12q subtelomeric sequences as bona fide TERRA promoters consisted of the silencing of these regions through a sgRNA-guided CRISPR-interference (CRISPRi) system. The CRISPRi system can modulate gene expression without altering the target DNA sequences. It consists of a catalytically inactive Cas9 (dCas9) fused with a transcriptional repressor domain known as the Kruppel-associated box (KRAB), which silence the target loci [99]. dCas9 strongly binds to the DNA target sequence and this tight binding interferes with the activity of transcription factors and RNAPII. KRAB works by recruiting additional co-repressor proteins such as KRAB-box-associated protein-1 (KAP-1) and epigenetic readers such as heterochromatin protein 1 (HP1), thus performing epigenetic reprograming of histone modifications and repressing RNAPII [100]. Fusing the KRAB repressor to dCas9 results in a stronger and more specific gene repressor than dCas9 alone [100].

U2OS cells were transfected with a dCas9-KRAB plasmid containing the sgRNA complementary to 7q (CRISPRi_7q U2OS) or 12q (CRISPRi_12q U2OS) sequences upstream of the TERRA TSS, or with an empty dCas9-KRAB vector (CRISPRi-EV U2OS). The results of the RT-qPCR with RNA extracted from the three U2OS populations can be seen below (Figure 24). If the targeted 7q and 12q subtelomeric sequences were indeed TERRA promoters one should observe a decrease in TERRA transcribed from these telomeres upon CRISPRi-derived silencing.

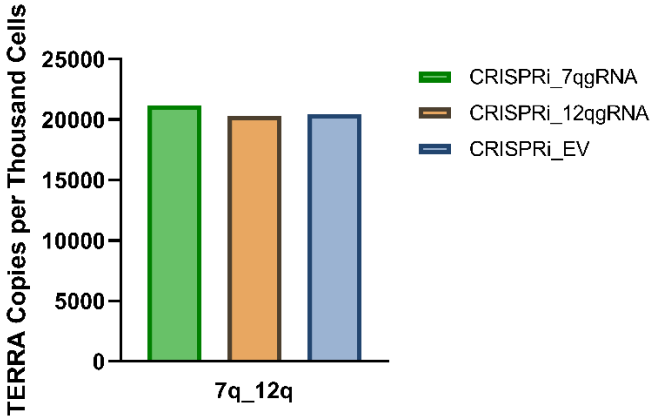

Figure 24: qPCR quantification of TERRA from chromosomes 7q and 12q in CRISPRi_7q, CRISPRi_12q and CRISPRi_EV U2OS cells. Reverse transcription performed with primer o1 (C-rich RT primer from Table 2). Absolute quantification. TERRA Copies per Thousand Cells displayed on y axis.

Approximately the same TERRA levels were measured in CRISPRi_7q U2OS and CRISPRi_12q U2OS as in the negative control CRISPRi-EV U2OS. It is possible that this is evidence for low activity of this sequence as a promoter in vivo. However, that does not match the ONT-seq results showing TERRA transcript coverage in this region, nor the RT-qPCR validation. It is more likely that there are limitations in the CRISPRi silencing protocol. This CRISPRi system and protocol have been used successfully for gene depletion in the Azzalin lab before. However, in the current experiment the designed sgRNAs annealed to the DNA sequence approximately 50 bp (for 7q) and 30 bp (for 12q) upstream of the TERRA TSS suggested by the ONT-seq analysis. An sgRNA that anneals on top of or immediately downstream of the TSS is preferred in these assays. It is possible that the dCas9 protein

is interacting with the DNA too far upstream from where TERRA transcription begins, and thus not interfering with TERRA transcription. However, due to the requirements of gRNA design it was not possible to design sgRNAs that annealed in the ideal locations and had high efficacy and specificity scores. This is further discussed in Section 4.5

4. Discussion and Future Work

In this project, TERRA in human cancer cells was quantified by RT-qPCR with the highest level of accuracy that current approaches can provide [18]. The results presented in this thesis, together with the long-read TERRA sequencing results obtained in the Azzalin lab using ONT, paint one of the clearest pictures so far of the TERRA transcription landscape in human cancer cells.

The data reinforce that the number of TERRA molecules originating from different chromosome ends can vary drastically. It also confirms, apart from irregularities in the SAOS2 cell line results, that ALT cancer cells express higher levels of TERRA. The measured number of TERRA molecules per cell varies from an average of less than one per cell in telomerase-positive cell lines to 10-20 molecules per cell in U2OS.

4.1. Validation of 7q_12q TERRA Quantification as Diagnostic Tool

The ONT-seq data reported that ALT cancer cells had higher abundance of TERRA transcripts than telomerase-positive cancer cells. However, it suggested that this increase was considerably more pronounced at chromosome ends 7q and 12 than at any other chromosome end. This led to an investigation of the potential of 7q- and 12q- targeted TERRA quantification for ALT cancer diagnosis.

The hypothesis for a 7q_12q TERRA quantification diagnosis rested on establishing a 7q_12q TERRA value cut-off. When analyzing a patient RNA sample, 7q_12q TERRA values above the cut-off would be indicative of an ALT cancer, while values below the cut-off would exclude it. For this test to have diagnostic accuracy (i. e. high discriminative potential), its sensitivity and specificity would have to be high, in order to avoid false positives (non-ALT cancers with 7q_12q TERRA levels above the cutoff) or false negatives (ALT cancers with 7q_12q TERRA levels below the cutoff) [101]. For this, ALT 7q_12q TERRA expression would ideally be orders of magnitude higher than telomerase-positive 7q_12q TERRA expression. This way, a threshold value could be established so that, even with high variation between samples, there would be low probability of an ALT cancer falling below the cut-off or a telomerase-positive cancer to score above it. Specifically, this study hypothesized that a minimum ALT 7q_12q TERRA fold-increase 50-to-100-fold relative to telomerase-positive 7q_12q TERRA expression could be sufficient for establishing a diagnosis threshold.

Observing the results, U2OS ALT cells had a higher average TERRA expression than HOS telomerase-positive cell lines, by both relative and absolute quantification. However, the maximum fold increase was only of approximately 13-fold and often less than that. Adding to this, a high variability between samples of the same cell line, especially in ALT cell lines, was observed. Indeed, the lowest reported levels of U2OS 7q_12q TERRA levels are only approximately 40% higher than the highest reported HOS 7q_12q TERRA levels. Therefore, for these two cell lines, it would not be possible to establish a TERRA level cut-off that could confidently segregate ALT and non-ALT cancers based on 7q_12q TERRA expression. This indicates that 7q_12q TERRA quantification

through an RT-qPCR assay as designed in this project is not viable as an ALT cancer diagnosis. Despite this, a wider panel of ALT and telomerase-positive cell lines will be tested in the following section of the project.

The high variability in detected 7q_12q TERRA transcripts between samples can be explained by two factors. The first is the possibility of naturally high fluctuations of TERRA in ALT cells. TERRA has been shown to increase in response to DNA damage induced by TRF2 depletion, and is believed to help sustain the telomeric DNA damage response [102]. ALT cell telomeres are, by nature, unstable, having to maintain a minimum basal level of replicative stress and DNA damage necessary for ALT to occur. This DNA damage level fluctuates under action of a complex network of activators and repressor, and TERRA may fluctuate in response to these alterations. It is plausible that moderate TERRA fluctuations (both total and 7q_12q specific) can occur in ALT populations in response to even minor stresses that the population is exposed to.

Another reason for variability between samples are technical factors in the RT-qPCR experiment. The reverse transcription process of each of the RNA is assumed to have identical efficiency for quantification purposes, but minor changes in RT efficiency from sample to sample might occur, which can add to a significant error when dealing with an RNA expressed at low levels, such as TERRA. Furthermore, RNA extraction yield is variable, so there are inequalities in RNA amounts isolated from equal cell numbers. Therefore, this RT-qPCR assay is difficult to fully standardize and is expected to have moderate standard deviations in a small series of samples.

TERRA expression variability can also be caused by a difference in the proliferative state of the used cells. This can be affected by their confluency at the time of extraction, and age of the population. Cells that slow their proliferation might decrease their ALT activity, and therefore TERRA expression. This factor is unlikely to have caused large variations in TERRA levels between samples in this project, since a considerable effort was put into assuring the standardization of cell confluency at the time of extraction.

However, in a clinical scenario it would be hard to assure the same standardization in samples obtained from patients as they could have been exposed to different stresses and conditions, and a moderate-to-high variability is to be expected, further cementing the necessity of a high order of magnitude of difference in 7q_12q TERRA levels.

As a last point, it is worth noting that quantification of total TERRA, rather than chromosome end specific TERRA, could potentially be a viable diagnosis tool. This could be achieved by RT-qPCR combined with rolling circle amplification (RCA) using padlock probes. A padlock probe is an oligonucleotide that, via a ligation reaction, forms circular DNA when hybridizing to specific target DNA. The region of the padlock probe that does not participate in target DNA hybridization contains generic primer sequences for amplification [103]. Padlock probes can be designed to target telomeric repeats in cDNA reverse transcribed from RNA sample. Circularized padlock probes are amplified by RCA, and then, real time qPCR is performed, allowing for quantitative assessment of total telomeric repeats in the RNA sample. This results in accurate estimation of TERRA transcripts. It is possible that, unlike in the case of chromosome end-specific TERRA, the fold-difference in total TERRA repeats could be as high as 50-to-100-fold, and allow for establishment of a diagnosis threshold.

4.2. Validation of ONT-seq by RT-qPCR

A broader validation of the ONT-seq results was performed for a panel of 4 cell lines, U2OS and SAOS2 ALT cells and HeLa and HEK293T telomerase-positive cells, at 5 chromosome ends, 5p, 7q_12q,15q,16p and 20q (Figure 15).

Higher TERRA levels were detected at 5p and 16p chromosome ends in both ALT and telomerase-positive cells than predicted by ONT-seq. The ONT analysis predicted very low TERRA expression for all cell lines at 5p and only moderate expression at 16p, yet both ends seemed to express more than the 7q_12q in the RT-qPCR results.

However, the standard curves used to calibrate the 5p and 16p results were quite unproportioned compared to the 7q_12q, 15q and 20q curves. The R^2 value associated with these 4-point curves was between 0.9 and 0.93, values that are too low to provide a good confidence within the correlation. The slopes of these curves are steeper than the slopes obtained for the other chromosome ends, using the same 1:10 dilution series, and they are likely to have caused an overestimation of TERRA levels. This is caused by an inability of the primers to recognize different amounts of the 5p and 16p target sequences with the same efficiency. Two pairs of forward and reverse primers were designed for both the 5p and 16p chromosome ends, however, none of them had R^2 >0.99 in primer efficiency testing. Due to the short length and repetitiveness of the target sequence, ideal primers could not be designed, and the assays were performed using the best possible primers.

This displays one of the limitations of the RT-qPCR assay, that it is highly dependent on the quality of the primers for the target sequence. Without efficient primers it is difficult to obtain reliable quantitative values and a true comparison between different ends is only possible when primers with identical efficiencies can be designed. Designing primers specifically for the subtelomeric regions is difficult due to the regions' high CG content, repetitive nature and similarity between chromosome ends. Furthermore, the target region for primers, between TERRA TSSs and the first telomeric 5'-TTAGGG-3' repeats, are very short length, which means there are few possibilities for primer design. Among the chromosome ends chosen for this study, 5p and 16p had particularly high repetitiveness and high GC content. Further attempts to design more effective primers for these ends can be performed in the future. Another limitation of the RT-qPCR assay comes from the possibility that TERRA transcription at some chromosome ends starts so close to the telomeric repeats that detection by chromosome end specific RT-PCR is made impossible [18].

In 7q_12q, 15q and 20q chromosome ends, RT-qPCR results validated most of the ONT-seq results. 20q end was confirmed to express low levels of TERRA, which is consistent with this subtelomere lacking the 29-37 bp promoter [16]. 15q expressed lower levels of TERRA than 7q_12q, as suggested by the ONT-seq results as well.

The only cell line that was not validated by RT-qPCR in these chromosome ends was the SAOS2 osteosarcoma cell line. ONT-seq detected high TERRA levels, particularly in 7q_12q, which is consistent with this being an ALT cell line. However, RT-qPCR detected low TERRA expression, comparable with the telomerase-positive cell lines. Several possibilities for this discrepancy were tested and ruled out. Primer annealing efficiency was proved to be similar in all cell lines via a control qPCR with known quantities of genomic DNA. Telomeric repeat degeneration

in SAOS2 was ruled out, as the results of the RT-qPCR were similar when using cDNA from a reverse transcription made with a subtelomeric sequence primer. And secondary structures in the SAOS2 sequence biasing reverse transcription efficiency between the 2 assays were ruled out by using matching temperatures for the reverse transcription reaction.

As explained in Section 4.1, TERRA levels fluctuate in cells in response to DNA damage levels. It is possible that this SAOS2 population simply suffered changes in the network of transcription regulators that resulted in a downregulating of TERRA expression. It is possible that there are limitation in the ONT-seq results as well. Before the sequencing, an RNA purification process was performed to produce TERRA-enriched RNA. A TERRA-enriched sample is less complex than a sample of total RNA. Less complex samples will likely have a more efficient cDNA synthesis, because there is less interference of other molecules in the process. In the ONT-seq experiment, TERRA was also purified from nuclear RNA, while in the RT-qPCR, it was purified from total RNA. These differences in the ONT-seq process could have resulted in a biased detection of 7q_12q TERRA in SAOS2 cells. Alternatively, the normalization and smoothening process during bioinformation analysis of the sequencing data could have caused a bias. In depth troubleshooting of ONT-seq data is beyond the scope of this thesis, however, in general, one of the greatest advantages of RT-qPCR over RNA-seq assays is that it is a more controlled and directed assay.

There is, however, a series of conclusions that can definitively be made from this set of RT-qPCR assays. First, TERRA was confirmed to be expressed from multiple chromosome ends in all analyzed human cell lines. This falls in line with the popular TERRA transcription model that postulates TERRA loci at multiple telomeres [18][86].

Finally, the 7q_12q conserved sequence was confirmed to be a considerable source of TERRA transcripts. However, these chromosome ends do not contain the previously characterized 29-37 bp promoters, that guide active TERRA transcription from other chromosome ends [16]. This led to the establishment of the hypothesis that a previously uncharacterized TERRA promoter in this conserved sequence might be actively transcribing TERRA in vivo. Thus, the next stage of this project consisted of a characterization of the 7q and 12q conserved sequence promoter activity. Furthermore, the effects of hypomethylation of this sequence on TERRA transcription was also studied.

4.3. 7q_12q TERRA Transcription Regulation by DNA Methyltransferases

Nergadze et al. showed that 29-37 repeats are methylated in vivo, and cytosine methylation negatively regulates TERRA abundance [16]. They found that concomitant disruption of DNMT1 and DNMT3b genes completely abolished methylation at 29-37 repeat CpG dinucleotide. Northern blot analysis showed a vast increase of TERRA levels in DKO cells. In a further study, HCT116 DKO cells were again shown to overexpress TERRA in relation to parental HCT116 in various chromosome ends containing subtelomeric 29-37 repeats (9p, 10q, 12q, 15q and, XqYq) (61-29-37 repeats distribution can be seen in Figure 5). HCT116 DKO cells also overexpressed TERRA from

several chromosome ends with no reported 61-29-37 repeats, but with considerably sized subtelomeric CpG islands, including the 12q chromosome end. [18]

Since DNA methylation regulates TERRA expression from 29-37 repeats promoters, this project investigated to what extent it could have a similar regulatory effect at the 7q and 12q sequence. Unlike previous studies, this project also aimed to perform absolute quantification of the DNA methylation regulatory effect by measuring the increase in 7q_12q TERRA copies per cell.

However, chromosome specific RT-qPCR detected similar 7q_12q TERRA expression in HCT116 and HCT116 DKO cells. To validate this result, RT-qPCR was performed in a positive control, 61-29-37 containing XqYq chromosome end [16], in which HCT116 DKO TERRA overexpression was reported by Linger et al. [18]. TERRA expression in this control was also identical, revealing an abnormal behavior in these populations of cell lines and resulting in the experiment having inconclusive results.

This assay should be repeated with new populations of HCT116 and DKO cells, or alternatively, with different cell lines knocked out for methyltransferases, to confirm the effects of subtelomeric DNA methylation on 7q_12q transcription.

DNA methylation is only the first of a series of regulatory factors that can be explored in the conserved 7q and 12q sequence. A search for transcription binding factors in the regions upstream of the 7q- and 12- TERRA TSSs has been performed in the Azzalin lab. This search did not find motifs of the TERRA transcription regulator CTCF in the 6000 bp upstream of the TSS, but a variety of other transcription binding factor sites were located (Supplementary Table A 4 many of which could be regulating active TERRA transcription at these ends. Further research into these transcription binding factors is still required. Assays can be performed by knocking down selected transcription factors followed by analysis of TERRA expression. Finding transcription factors and epigenetic modifications that are driving active TERRA transcription at these chromosome ends would be important to understand the contribution of 7q and 12q subtelomeres to TERRA profiles in human cancer cells.

4.4. 7q and 12q Putative Promoter Isolation and Future Work

The 7q and 12q CpG islands were identified using two independent bioinformatic tools, MethPrimer 2.0 and the CpGPlot/CpGReport, and were amplified by PCR using chromosome end-specific primers. The similarity between the 7q and 12q chromosome ends is such that it is difficult to design primers that distinguish one from another. Hence, the two sequences were amplified using identical primers. The smaller 12q CpG island is very similar to a subsection of the larger 7q CpG island, except for the presence of an extra 30 bp sequence. These 30 extra bp were used to distinguish between the two sequences following amplification. Then, they were inserted into a pJET1.2 plasmid. The isolation of these sequences paves the way for assays to functionally characterize the putative TERRA promoter within the CpG islands.

To test the sequences for promoter activity, a promoter reporter assay can be developed. For this, the CpG island sequences can be excised from the pJET1.2 vector with restriction enzymes and cloned into a reporter plasmid upstream of a reporter gene such as green fluorescence protein (eGFP). eGFP fluorescence can be quantified by densitometric analysis. Furthermore, eGFP mRNA can be quantified by reverse transcribing and performing RT-qPCR. Similarly, fluorescence microscopy analysis of human cells transfected with the CpG-containing reporter plasmids would reveal if the inserted sequences induce cellular eGFP expression, indicating that they have promoter activity. Cloning the sequences in both directions upstream of the eGFP would also allow for testing promoter directionality. This is relevant as the 29bp repeat of the 61-29-37 bp repeats promoter is known to start transcription in both directions [16], and the same could be true for the 7q_12q putative promoter. Through this method, the putative promoter strength can be estimated by comparison with other promoters such as the strong cytomegalovirus (CMV) promoter and the 29-37 bp promoter.

In addition, future work can still be done to investigate the 7q and 12q putative promoters. For example, investigating RNAPII enrichment in vivo at these sequences by chromatin immunoprecipitation experiments. These can be performed on chromatin from human ALT cells using independent antibodies against phosphorylation-activated RNAPII large subunit.

Another assay could involve monitoring of ATRS and DNA damage at telomeres following knockout of the 7q and 12q sequences. DNA damage markers can be used, such as phosphorylated histone H2AX (γH2AX), which forms µm-sized nuclear foci at the sites of DNA DSBs. The typical DNA damage monitoring assay would consist of immunofluorescence staining of cells knocked out for 7q and 12q sequences with antibodies against γH2AX and a telomeric marker such as TRF2.

Finding a new TERRA promoter in these sequences would expand the current catalogue of known TERRA promoters and contribute a better mapping of TERRA loci and compiling of the TERRA transcription profile in human cancer cells.

4.5. 7q_12q TERRA Depletion by CRISPRi Assay

The final assay performed in this project was done to downregulate 7q_12q TERRA expression in U2OS cancer cells using a CRISPRi system to silence the putative CpG island promoters using a dCas9-KRAB plasmid guided by sgRNAs to the putative promoters. If the sequences promoted TERRA transcription, then the result of this assay was expected to be depletion of 7q_12q TERRA, measurable by RT-qPCR,

However, 7q_12q TERRA levels were equal in U2OS cells transfected with dCas9-KRAB plasmids guided with sgRNAs towards 7q or 12q, and in negative control cells transfected with dCas9-KRAB with no sgRNA. This could be biologically interpreted as a lack of promoter activity in vivo, but is more likely a result of experimental limitations. Most notably, the designed sgRNAs annealed to the DNA sequence upstream of the TERRA

transcription start site, and is plausible that the dCas9-KRAB complex is targeting a sequence too far upstream to interfere with RNAPII binding and transcriptional elongation.

It was not possible to design sgRNAs directed towards an ideal target, which is immediately downstream of the TERRA TSS, as the sgRNA require a very specific design to function efficiently. Briefly, the CRISPRi system mimics a system from *Streptococcus pyogenes* that relies Cas9 protein and two RNAs, a mature CRISPR RNA (crRNA) and a partially complementary trans-acting RNA (tracrRNA). The sgRNAs designed for the CRISPRi must include a hairpin that mimics the tracrRNA-crRNA complex [104]. sgRNA binding specificity to the DNA target is also dependent on the presence of a short DNA motif (protospacer adjacent motif [PAM] sequence: NGG) juxtaposed to the DNA complementary region. The web tool used to design sgRNAs in this study, the CRISPR-ERA webtool, locates sequences in a target DNA that meet the criteria to be targets for sgRNA binding, and assigns them scores based on efficacy and specificity (E+S). There were no outputs with high E + S scores in the 7q and 12q CpG sequences that were located immediately downstream of the TERRA TSS. Therefore, it was not possible to design the CRISPRi system in its ideal form.

5. Conclusion

This project was designed to increase our understanding of TERRA transcription profiles in human cancer cells. Through extensive and accurate RT-qPCR measurement of TERRA, validating previous long-read RNA-seq results, TERRA transcription was studied in a panel of ALT and telomerase-positive cancer cells at multiple chromosome ends.

TERRA was confirmed to be expressed from multiple chromosome ends in all analyzed human cell lines, supporting the notion that TERRA originates from several telomeres. Chromosome-specific TERRA quantification was postulated to become a potential ALT diagnosis procedure. However, expression was shown to be prone to fluctuations and the fold-increase in ALT cells is likely not sufficient for a robust diagnosis method to be established.

Elevated expression of TERRA was found in U2OS ALT cells at the long arms of 7q and 12q chromosome ends, leading to the establishment of the hypothesis that these ends contain a previously uncharacterized promoter actively transcribing TERRA. The subtelomeric CpG island sequence containing a TERRA transcription start site, which is conserved between the two ends, was isolated. The first steps were taken to characterize its promoter activity and further assays will be performed in the Azzalin lab with this goal.

This project helped gain new insights into the profile of TERRA transcription in ALT cells and helped guide an ongoing search into further sites of active TERRA transcription. The identified TERRA loci could represent potential targets for TERRA-focused approaches. Research into this area is of high importance, since ALT cancers reliant on TERRA transcription currently affect millions of people, have generally poor prognosis and are resistant to currently available treatments.

6. Bibliography

[1] N. Hug and J. Lingner, "Telomere length homeostasis," *Chromosoma*, pp. 413–425, 2006, doi: 10.1007/s00412-006-0067-3.

[2] A. Prieur and D. S. Peeper, "Cellular senescence in vivo: a barrier to tumorigenesis," *Curr. Opin. Cell Biol.*, vol. 20, no. 2, pp. 150–155, 2008, doi: 10.1016/j.ceb.2008.01.007.

[3] Y. Deng, S. S. Chan, and S. Chang, "Telomere dysfunction and tumour suppression: The senescence connection," *Nat. Rev. Cancer*, vol. 8, no. 6, pp. 450–458, 2008, doi: 10.1038/nrc2393.

[4] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011, doi: 10.1016/j.cell.2011.02.013.

[5] M. A. Feitelson *et al.*, "Sustained proliferation in cancer: Mechanisms and novel therapeutic targets," *Semin. Cancer Biol.*, vol. 35, pp. S25–S54, 2015, doi: 10.1016/j.semcancer.2015.02.006.

[6] B. P. O. DeRisi Joseph, Penland Lolita, S. Tyagi, F. R. Kramer, N. P. Group, and B. P. O. DeRisi Joseph, Penland Lolita, "Evidence for an Alternative Mechanism for Maintaining Telomere Length in Human Tumors and Tumor-Derived Cell Lines," *Nat. Puplishing Gr.*, vol. 4, pp. 303–308, 1996, doi: 10.1038/nm0798-822.

[7] J. D. Henson, A. A. Neumann, T. R. Yeager, and R. R. Reddel, "Alternative lengthening of telomeres in mammalian cells," *Oncogene*, vol. 21, no. 4 REV. ISS. 1, pp. 598–610, 2002, doi: 10.1038/sj/onc/1205058.

[8] A. J. Cesare and R. R. Reddel, "Alternative lengthening of telomeres: Models, mechanisms and implications," *Nat. Rev. Genet.*, vol. 11, no. 5, pp. 319–330, 2010, doi: 10.1038/nrg2763.

[9] C. M. Heaphy *et al.*, "Prevalence of the Alternative Lengthening of Telomeres Telomere Maintenance Mechanism in Human Cancer Subtypes," *Short Commun.*, vol. 179, no. 4, pp. 1608–1615, 2011, doi: 10.1016/j.ajpath.2011.06.018.

[10] R. T. Lawlor *et al.*, "Alternative lengthening of telomeres (ALT) influences survival in soft tissue sarcomas : a systematic review with meta-analysis," *BMC Cancer*, pp. 1–7, 2019.

[11] J. M. Zhang and L. Zou, "Alternative lengthening of telomeres: From molecular mechanisms to therapeutic outlooks," *Cell Biosci.*, vol. 10, no. 1, pp. 1–9, 2020, doi: 10.1186/s13578-020-00391-6.

[12] C. M. Azzalin, P. Reichenbach, L. Khoriauli, E. Giulotto, and J. Lingner, "Telomeric Repeat–Containing RNA and RNA Surveillance Factors at," *Science (80-.).*, vol. 318, no. November, pp. 798–801, 2007.

[13] C. A. Lovejoy *et al.*, "Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of Telomeres pathway," *PLoS Genet.*, vol. 8, no. 7, pp. 12–15, 2012, doi: 10.1371/journal.pgen.1002772.

[14] R. Arora, Y. Lee, H. Wischnewski, C. M. Brun, T. Schwarz, and C. M. Azzalin, "RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells," *Nat. Commun.*, vol. 5, pp. 1–11, 2014, doi: 10.1038/ncomms6220.

[15] A. Porro, S. Feuerhahn, J. Delafontaine, H. Riethman, J. Rougemont, and J. Lingner, "Functional characterization of the TERRA transcriptome at damaged telomeres," *Nat. Commun.*, vol. 5, pp. 1–13, 2014, doi: 10.1038/ncomms6379.

[16] S. G. Nergadze *et al.*, "CpG-island promoters drive transcription of human telomeres," *Rna*, vol. 15, no. 12, pp. 2186–2194, 2009, doi: 10.1261/rna.1748309.

[17] D. Montero, Silanes, O. Gran, and M. A. Blasco, "Telomeric RNAs are essential to maintain telomeres," *Nat. Commun.*, pp. 1–13, 2016, doi: 10.1038/ncomms12534.

[18] M. Feretzaki, P. R. Nunes, and J. Lingner, "Expression and differential regulation of human TERRA at

several chromosome ends," *Rna*, vol. 25, no. 11, pp. 1470–1480, 2019, doi: 10.1261/rna.072322.119.

[19] W. Palm and T. De Lange, "How shelterin protects mammalian telomeres," *Annu. Rev. Genet.*, vol. 42, pp. 301–334, 2008, doi: 10.1146/annurev.genet.41.110306.130350.

[20] R. J. O'Sullivan and J. Karlseder, "Telomeres: Protecting chromosomes against genome instability," *Nat. Rev. Mol. Cell Biol.*, vol. 11, no. 3, pp. 171–181, 2010, doi: 10.1038/nrm2848.

[21] Titia De Lange, "How Telomeres Solve the End-Protection Problem," *Science (80-.).*, vol. 948, no. 2009, 2009, doi: 10.1126/science.1170633.

[22] R. K. Moyzis *et al.*, "A highly conserved repetitive DNA sequence, (TTAGGG)., present at the telomeres of human chromosomes," *Proc. Natl. Acad. Sci. USA*, vol. 85, no. September, pp. 6622–6626, 1988.

[23] W. Lu, Y. Zhang, D. Liu, Z. Songyang, and M. Wan, "Telomeres — structure , function , and regulation," *Exp. Cell Res.*, vol. 319, no. 2, pp. 133–141, 2013, doi: 10.1016/j.yexcr.2012.09.005.

[24] J. Kanoh and F. Ishikawa, "Cellular and Molecular Life Sciences Composition and conservation of the telomeric complex," *CMLS*, vol. 60, pp. 2295–2302, 2003, doi: 10.1007/s00018-003-3245-y.

[25] E. Cusanelli and P. Chartrand, "Telomeric repeat-containing RNA TERRA: A noncoding RNA connecting telomere biology to genome integrity," *Front. Genet.*, vol. 6, no. MAR, pp. 1–9, 2015, doi: 10.3389/fgene.2015.00143.

[26] T. De Lange, "Shelterin: The protein complex that shapes and safeguards human telomeres," *Genes Dev.*, vol. 19, no. 18, pp. 2100–2110, 2005, doi: 10.1101/gad.1346005.

[27] A. Sfeir and T. De Lange, "Removal of shelterin reveals the telomere end-protection problem," *Science (80-.).*, vol. 336, no. 6081, pp. 593–597, 2012, doi: 10.1126/science.1218498.

[28] Y. Doksani, J. Y. Wu, T. De Lange, and X. Zhuang, "XSuper-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation," *Cell*, vol. 155, no. 2, p. 345, 2013, doi: 10.1016/j.cell.2013.09.048.

[29] T. de Lange, "Shelterin-Mediated Telomere Protection," *Annu. Rev. Genet.*, vol. 52, no. 1, pp. 223–247, 2018, doi: 10.1146/annurev-genet-032918-021921.

[30] N. Bennardo, A. Cheng, N. Huang, and J. M. Stark, "Alternative-NHEJ Is a Mechanistically Distinct Pathway of Mammalian Chromosome Break Repair," *PLoS Genet.*, vol. 4, no. 6, 2008, doi: 10.1371/journal.pgen.1000110.

[31] E. L. Denchi and T. De Lange, "Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1," *Nature*, vol. 448, no. 7157, pp. 1068–1071, 2007, doi: 10.1038/nature06065.

[32] D. Hockemeyer *et al.*, "Telomere protection by mammalian Pot1 requires interaction with Tpp1," *Nat. Struct. Mol. Biol.*, vol. 14, no. 8, pp. 754–761, 2007, doi: 10.1038/nsmb1270.

[33] I. Schmutz, L. Timashev, W. Xie, D. J. Patel, and T. De Lange, "TRF2 binds branched DNA to safeguard telomere integrity," *Nat. Struct. Mol. Biol.*, vol. 24, no. 9, pp. 734–742, 2017, doi: 10.1038/nsmb.3451.

[34] K. K. Takai, T. Kibe, J. R. Donigian, D. Frescas, and T. de Lange, "Telomere Protection by TPP1/POT1 Requires Tethering to TIN2," *Mol. Cell*, vol. 44, no. 4, pp. 647–659, 2011, doi: 10.1016/j.molcel.2011.08.043.

[35] J. Sarthy, N. S. Bae, J. Scrafford, and P. Baumann, "Human RAP1 inhibits non-homologous end joining at telomeres," *EMBO J.*, vol. 28, no. 21, pp. 3390–3399, 2009, doi: 10.1038/emboj.2009.275.

[36] D. Hockemeyer, J. P. Daniels, H. Takai, and T. de Lange, "Recent Expansion of the Telomeric Complex in Rodents: Two Distinct POT1 Proteins Protect Mouse Telomeres," *Cell*, vol. 126, no. 1, pp. 63–77, 2006, doi: 10.1016/j.cell.2006.04.044.

[37] R. C. Wang, A. Smogorzewska, and T. De Lange, "Homologous Recombination Generates T-Loop-Sized Deletions at Human Telomeres," *Cell*, vol. 119, pp. 355–368, 2004, [Online]. Available: http://www.

[38] W. Palm, D. Hockemeyer, T. Kibe, and T. de Lange, “Functional Dissection of Human and Mouse POT1 Proteins,” *Mol. Cell. Biol.*, vol. 29, no. 2, pp. 471–482, 2009, doi: 10.1128/mcb.01352-08.

[39] G. B. Celli and T. de Lange, “DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion,” *Nat. Cell Biol.*, vol. 7, no. 7, pp. 712–718, 2005, doi: 10.1038/ncb1275.

[40] M. Z. Levy, R. C. Allsopp, A. B. Futcher, C. W. Greider, and C. B. Harley, “Telomere end-replication problem and cell aging,” *J. Mol. Biol.*, vol. 225, no. 4, pp. 951–960, 1992, doi: 10.1016/0022-2836(92)90096-3.

[41] V. Lundblad and J. W. Szostak, “A mutant with a defect in telomere elongation leads to senescence in yeast,” *Cell*, vol. 57, no. 4, pp. 633–643, 1989, doi: 10.1016/0092-8674(89)90132-3.

[42] M. A. Blasco, “Replicating through telomeres : a means to an end,” *CellPress*, pp. 1–12, 2015, doi: 10.1016/j.tibs.2015.06.003.

[43] A. Smogorzewska and T. De Lange, “Regulation of telomerase by telomeric proteins,” *Annu. Rev. Biochem.*, vol. 73, pp. 177–208, 2004, doi: 10.1146/annurev.biochem.73.071403.160049.

[44] H. Niida *et al.*, “Severe growth defect in mouse cells lacking the telomerase RNA component,” *Nat. Genet.*, vol. 19, no. 2, pp. 203–206, 1998, doi: 10.1038/580.

[45] M. A. Blasco *et al.*, “Telomere shortening and tumor formation by mouse cells lacking telomerase RNA,” *Cell*, vol. 91, no. 1, pp. 25–34, 1997, doi: 10.1016/S0092-8674(01)80006-4.

[46] C. B. Harley, A. B. Futcher, and C. W. Greider, “Telomeres shorten during ageing of human fibroblasts,” *Nature*, vol. 345, no. 6274, pp. 458–460, 1990, doi: 10.1038/345458a0.

[47] C. M. Azzalin and J. Lingner, “Telomeres: The silence is broken,” *Cell Cycle*, vol. 7, no. 9, pp. 1161–1165, 2008, doi: 10.4161/cc.7.9.5836.

[48] C. B. Harley, “Telomere loss: mitotic clock or genetic time bomb?,” *Mutat. Res. DNAging*, vol. 256, no. 2–6, pp. 271–282, 1991, doi: 10.1016/0921-8734(91)90018-7.

[49] S. A. Stewart and R. A. Weinberg, “Telomeres: Cancer to Human Aging,” *Annu. Rev. Cell Dev. Biol.*, vol. 22, no. 1, pp. 531–557, 2006, doi: 10.1146/annurev.cellbio.22.010305.104518.

[50] T. Finkel, M. Serrano, and M. A. Blasco, “The common biology of cancer and ageing,” *Nature*, vol. 448, no. 7155, pp. 767–774, 2007, doi: 10.1038/nature05985.

[51] C. M. Azzalin and J. Lingner, “Telomere functions grounding on TERRA firma,” *Trends Cell Biol.*, vol. 25, no. 1, pp. 29–36, 2015, doi: 10.1016/j.tcb.2014.08.007.

[52] N. W. Kim *et al.*, “Specific Association of Human Telomerase Activity with Immortal Cells and Cancer Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2885288 Accessed : 18-04-2016 15 : 00 UTC Your use of the JSTOR ar,” *Science (80-.).*, vol. 266, no. 5193, pp. 2011–2015, 1994.

[53] C. W. Greider, “Telomeres, Telomerase and Senescence,” *BioEssays*, no. 1, pp. 66–71, 1990.

[54] I. Draskovic and A. L. Vallejo, “Telomere recombination and alternative telomere lengthening mechanisms,” *Front. Biosci.*, vol. 18, no. 1, pp. 1–20, 2013, doi: 10.2741/4084.

[55] C. W. Greider and E. H. Blackburn, “Identification of a specific telomere terminal transferase activity in tetrahymena extracts,” *Cell*, vol. 43, no. 2 PART 1, pp. 405–413, 1985, doi: 10.1016/0092-8674(85)90170-9.

[56] T. M. Bryan, A. Englezou, J. Gupta, S. Bacchetti, and R. R. Reddel, “Telomere elongation in immortal human cells without detectable telomerase activity,” *EMBO J.*, vol. 14, no. 17, pp. 4240–4248, 1995, doi: 10.1002/j.1460-2075.1995.tb00098.x.

[57] J. M. Zhang, T. Yadav, J. Ouyang, L. Lan, and L. Zou, “Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways,” *Cell Rep.*, vol. 26, no. 4, pp. 955-968.e3, 2019, doi: 10.1016/j.celrep.2018.12.102.

[58] H. Ogino *et al.*, “Release of telomeric DNA from chromosomes in immortal human cells lacking telomerase activity,” *Biochem. Biophys. Res. Commun.*, vol. 248, no. 2, pp. 223–227, 1998, doi: 10.1006/bbrc.1998.8875.

[59] Y. Tokutake *et al.*, “Extra-chromosomal telomere repeat DNA in telomerase-negative immortalized cell lines,” *Biochem. Biophys. Res. Commun.*, vol. 247, no. 3, pp. 765–772, 1998, doi: 10.1006/bbrc.1998.8876.

[60] J. D. Henson *et al.*, “DNA C-circles are specific and quantifiable markers of alternative- lengthening-of-telomeres activity,” *Nat. Biotechnol.*, vol. 27, no. 12, pp. 1181–1185, 2009, doi: 10.1038/nbt.1587.

[61] J. D. Henson, L. M. Lau, S. Koch, N. Martin La Rotta, R. A. Dagg, and R. R. Reddel, “The C-Circle Assay for alternative-lengthening-of-telomeres activity.,” *Methods*, vol. 114, pp. 74–84, Feb. 2017, doi: 10.1016/j.ymeth.2016.08.016.

[62] S. M. Bailey, M. A. Brenneman, and E. H. Goodwin, “Frequent recombination in telomeric DNA may extend the proliferative life of telemerase-negative cells,” *Nucleic Acids Res.*, vol. 32, no. 12, pp. 3743–3751, 2004, doi: 10.1093/nar/gkh691.

[63] J. Kramara, B. Osia, and A. Malkova, “Break-Induced Replication: The Where, The Why, and The How.,” *Trends Genet.*, vol. 34, no. 7, pp. 518–531, Jul. 2018, doi: 10.1016/j.tig.2018.04.002.

[64] S. Le, J. K. Moore, J. E. Haber, and C. W. Greider, “RAD50 and RAD51 define two pathways that collaborate to maintain telomeres in the absence of telomerase.,” *Genetics*, vol. 152, no. 1, pp. 143–152, May 1999.

[65] Q. Chen, A. Ijpma, and C. W. Greider, “Two survivor pathways that allow growth in the absence of telomerase are generated by distinct telomere recombination events.,” *Mol. Cell. Biol.*, vol. 21, no. 5, pp. 1819–1827, Mar. 2001, doi: 10.1128/MCB.21.5.1819-1827.2001.

[66] B. Domingues-Silva, B. Silva, and C. M. Azzalin, “ALTernative Functions for Human FANCM at Telomeres,” *Front. Mol. Biosci.*, vol. 6, no. September, pp. 1–7, 2019, doi: 10.3389/fmolb.2019.00084.

[67] B. Balk *et al.*, “Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence,” *Nat. Struct. Mol. Biol.*, vol. 20, no. 10, pp. 1199–1205, 2013, doi: 10.1038/nsmb.2662.

[68] M. Tarsounas and M. Tijsterman, “Genomes and G-quadruplexes: for better or for worse.,” *J. Mol. Biol.*, vol. 425, no. 23, pp. 4782–4789, Nov. 2013, doi: 10.1016/j.jmb.2013.09.026.

[69] K. E. Cox, A. Maréchal, and R. L. Flynn, “SMARCAL1 Resolves Replication Stress at ALT Telomeres.,” *Cell Rep.*, vol. 14, no. 5, pp. 1032–1040, Feb. 2016, doi: 10.1016/j.celrep.2016.01.011.

[70] B. Silva *et al.*, “FANCM limits ALT activity by restricting telomeric replication stress induced by deregulated BLM and R-loops,” *Nat. Commun.*, vol. 10, no. 1, pp. 1–16, 2019, doi: 10.1038/s41467-019-10179-z.

[71] R. Lu *et al.*, “The FANCM-BLM-TOP3A-RMI complex suppresses alternative lengthening of telomeres (ALT),” *Nat. Commun.*, vol. 10, no. 1, p. 2252, 2019, doi: 10.1038/s41467-019-10180-6.

[72] J. A. Baur, Y. Zou, J. W. Shay, and W. E. Wright, “Telomere Position Effect in Human Cells Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/3084038 Linked references are available on JSTOR for this article : Telomere Position Effect in Human Cells,” *Science (80-.).*, vol. 292, no. 5524, pp. 2075–2077, 2016.

[73] G. Rudenko and L. H. Van der Ploeg, “Transcription of telomere repeats in protozoa.,” *EMBO J.*, vol. 8, no. 9, pp. 2633–2638, 1989, doi: 10.1002/j.1460-2075.1989.tb08403.x.

[74] S. Schoeftner and M. A. Blasco, “Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II,” *Nat. Cell Biol.*, vol. 10, no. 2, pp. 228–236, 2008, doi: 10.1038/ncb1685.

[75] A. Diman and A. Decottignies, “Genomic origin and nuclear localization of TERRA telomeric repeat-containing RNA: from Darkness to Dawn,” *FEBS J.*, vol. 285, no. 8, pp. 1389–1398, 2018, doi: 10.1111/febs.14363.

[76] M. Feretzaki and J. Lingner, “A practical qPCR approach to detect TERRA, the elusive telomeric repeat-containing RNA,” *Methods*, vol. 114, pp. 39–45, 2017, doi: 10.1016/j.ymeth.2016.08.004.

[77] W. R. A. Brown, P. J. MacKinnon, A. Villasanté, N. Spurr, V. J. Buckle, and M. J. Dobson, “Structure and polymorphism of human telomere-associated DNA,” *Cell*, vol. 63, no. 1, pp. 119–132, 1990, doi: 10.1016/0092-8674(90)90293-N.

[78] B. O. Farnung, E. Giulotto, and C. M. Azzalin, “Promoting transcription of chromosome ends,” *Transcription*, vol. 1, no. 3, pp. 140–143, 2010, doi: 10.4161/trns.1.3.13191.

[79] N. Bettin, C. Oss Pegorar, and E. Cusanelli, “The Emerging Roles of TERRA in Telomere Maintenance and Genome Stability,” *Cells*, vol. 8, no. 3, p. 246, 2019, doi: 10.3390/cells8030246.

[80] N. Arnoult, A. Van Beneden, and A. Decottignies, “Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1α,” *Nat. Struct. Mol. Biol.*, vol. 19, no. 9, pp. 948–956, 2012, doi: 10.1038/nsmb.2364.

[81] Z. Deng, J. Norseen, A. Wiedmer, H. Riethman, and P. M. Lieberman, “TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres.,” *Mol. Cell*, vol. 35, no. 4, pp. 403–413, Aug. 2009, doi: 10.1016/j.molcel.2009.06.025.

[82] S. Sagie *et al.*, “Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids,” *Nat. Commun.*, vol. 8, pp. 1–12, 2017, doi: 10.1038/ncomms14015.

[83] E. Cusanelli, C. A. P. Romero, and P. Chartrand, “Telomeric Noncoding RNA TERRA Is Induced by Telomere Shortening to Nucleate Telomerase Molecules at Short Telomeres,” *Mol. Cell*, vol. 51, no. 6, pp. 780–791, 2013, doi: 10.1016/j.molcel.2013.08.029.

[84] A. Aguilera and T. García-Muse, “R Loops: From Transcription Byproducts to Threats to Genome Stability,” *Mol. Cell*, vol. 46, no. 2, pp. 115–124, 2012, doi: 10.1016/j.molcel.2012.04.009.

[85] Z. Deng *et al.*, “A role for CTCF and cohesin in subtelomere and telomere end protection,” *EMBO J.*, pp. 1–14, 2012, doi: 10.1038/emboj.2012.266.

[86] K. Beishline, O. Vladimirova, S. Tutton, Z. Wang, Z. Deng, and P. M. Lieberman, “CTCF driven TERRA transcription facilitates completion of telomere DNA replication,” *Nat. Commun.*, pp. 1–10, doi: 10.1038/s41467-017-02212-w.

[87] M. Feretzaki, M. Pospisilova, R. Valador Fernandes, T. Lunardi, L. Krejci, and J. Lingner, “RAD51-dependent recruitment of TERRA lncRNA to telomeres through R-loops,” *Nature*, vol. 587, no. 7833, pp. 303–308, 2020, doi: 10.1038/s41586-020-2815-6.

[88] C. Soneson, Y. Yao, A. Bratus-Neuenschwander, A. Patrignani, M. D. Robinson, and S. Hussain, “A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes,” *Nat. Commun.*, vol. 10, no. 1, p. 3359, 2019, doi: 10.1038/s41467-019-11272-z.

[89] C. de Lannoy, D. de Ridder, and J. Risse, “The long reads ahead: de novo genome assembly using the MinION.,” *F1000Research*, vol. 6, p. 1083, 2017, doi: 10.12688/f1000research.12012.2.

[90] C. A. Heid, J. Stevens, K. J. Livak, and P. M. Williams, “Real time quantitative PCR,” *Genome Res.*, vol. 6, no. 10, pp. 986–994, 1996, doi: 10.1101/gr.6.10.986.

[91] J. S. Yuan, A. Reed, F. Chen, and C. N. Stewart, “Statistical analysis of real-time PCR data,” *BMC Bioinformatics*, vol. 7, pp. 1–12, 2006, doi: 10.1186/1471-2105-7-85.

[92] M. Arya, I. S. Shergill, M. Williamson, L. Gommersall, N. Arya, and H. R. H. Patel, “Basic principles of real-time quantitative PCR,” *Expert Rev. Mol. Diagn.*, vol. 5, no. 2, pp. 209–219, 2005, doi: 10.1586/14737159.5.2.209.

[93] M. W. Pfaffl, *Real-time PCR*. International University Line, 2007.

[94] P. Rice, L. Longden, and A. Bleasby, “EMBOSS: The European Molecular Biology Open Software Suite,” *Trends Genet.*, vol. 16, no. 6, pp. 276–277, 2000, doi: 10.1016/S0168-9525(00)02024-2.

[95] L. C. Li and R. Dahiya, “MethPrimer: Designing primers for methylation PCRs,” *Bioinformatics*, vol. 18, no. 11, pp. 1427–1431, 2002, doi: 10.1093/bioinformatics/18.11.1427.

[96] M. Lee *et al.*, “Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes,” *Nucleic Acids Res.*, vol. 42, no. 3, pp. 1733–1746, 2014, doi: 10.1093/nar/gkt1117.

[97] A. S. N. Alhendi and N. J. Royle, “The absence of (TCAGGG)n repeats in some telomeres, combined with variable responses to NR2F2 depletion, suggest that this nuclear receptor plays an indirect role in the alternative lengthening of telomeres,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, 2020, doi: 10.1038/s41598-020-77606-w.

[98] A. Sommer and N. J. Royle, “ALT: A multi-faceted phenomenon,” *Genes (Basel).*, vol. 11, no. 2, pp. 1–13, 2020, doi: 10.3390/genes11020133.

[99] K. M. Parsi, E. Hennessy, N. Kearns, and R. Maehr, “Using an Inducible CRISPR-dCas9-KRAB Effector System to Dissect Transcriptional Regulation in Human Embryonic Stem Cells.,” *Methods Mol. Biol.*, vol. 1507, pp. 221–233, 2017, doi: 10.1007/978-1-4939-6518-2_16.

[100] M. Adli, “The CRISPR tool kit for genome editing and beyond.,” *Nat. Commun.*, vol. 9, no. 1, p. 1911, May 2018, doi: 10.1038/s41467-018-04252-2.

[101] A. M. Simundic, “Measures of Diagnostic Accuracy: Basic Definitions,” *Ejifcc*, vol. 19, no. 4, pp. 203–211, 2009.

[102] A. Porro, S. Feuerhahn, and J. Lingner, “TERRA-Reinforced Association of LSD1 with MRE11 Promotes Processing of Uncapped Telomeres,” *Cell Rep.*, vol. 6, no. 4, pp. 765–776, 2014, doi: 10.1016/j.celrep.2014.01.022.

[103] J. Gong, Y. Li, T. Lin, X. Feng, and L. Chu, “Multiplex real-time PCR assay combined with rolling circle amplification (MPRP) using universal primers for non-invasive detection of tumor-related mutations,” *RSC Adv.*, vol. 8, no. 48, pp. 27375–27381, 2018, doi: 10.1039/C8RA05259J.

[104] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, “A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity,” *Science (80-.).*, vol. 337, no. August, pp. 816–822, 2012.

[105] D. M. Baird, A. J. Jeffreys, and N. J. Royle, “Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere.,” *EMBO J.*, vol. 14, no. 21, pp. 5433–5443, Nov. 1995.

Annex A

Table A 1: gBlock fragments used as standards for absolute quantification in this study

Oligo	Sequence (5’ -> 3´)	Length
5p	TATTGAGGCAGTGCATTAGCATACAGGTGCTTGTTACATGAGCAATGGGGGTGTCATATTTTGGGTG TCATGTCTGCATTAGGAATGCTGCATTTGTCTTCCGAGGCTGCGGTTGGATCTCGCACTGCAGCCCCG CCTTGCTTGGGAGAACCTCGGTGGGCAGGATTCAGAGGGGCTTTTGGTTTTCCTTTTCCACACTGAA CCCTTCTAACTGGTCTCTGACCCTGATTATTCAGGGCTGCAAACGGGAAGGATTTTATTCACCGTCTA TGCGGCCCCGAGTTGTCCCAAAGCGAGGCAGTGCCCCAAAGGTCTGTGCTGAGGAGAACGCTGCTC TGCCTTCGCGGTGTCCCCGGGTCTGTGCTGAGCAGAA	373
7q_12q	CACAGACCCGGGGGACACCGCGAAGGCAGAGCAGCGTTCTCCTCAGCACAGACCTTGGGGGCACTG CCTCGCTTTGGGACAACTCGGGGCCGCATCGACGGTGAATAAAATCCTTCCTGTTTGCAGCCCTGAA TAATCAGGGTCAGAGACCAGTTAGAAGGGTTCAGTGTGGAAAACGGGAAACCAAAAGCCCCTCTGA ATCCTGCCCACCGAGGTTCTCCCCAGCCAAGGCGAGGCGGCCGCAGTGCGAGATCCACACCGCAGC CTCGGAAGACAAATGCAGCATTCCTAATGCAGACATGACACCCAAAATATGACACACCCCATTGCTC ATGTAACAAGCACCTGTAATGCTAATGCACTGCCTCAATACAAAATATTAATATAAGATCCGCAATCC CCTTGCTGCCGTGCAGTCCTAAGACAGCGATCATAATAATCAACATTATA	450
15q	CTAACCCTAACCACATGAGCAACGTGGGTGTTATATTTTGGGTATCATGTGTGCATTAGGACTGCTGC ATTTGCGTTCCGATGCTGCAACTGGACCCTTTAATGCAGCCCCTCGCCTTAGCTTGGGAGAATCTCGC TGCCCAGGATTCAGAGGGGCTTTTGCTTTCCCGTTTTCCGCACTGAACCGCTCTAACTGGTCTCTGAC CTTGATTATTCACGGCTGCAACCGGGAAAGATTTTATTCACCGTTGATGCGCCCCGAGTTGTCCCAAA GCCACGCAGTGGCCCCAACGTCTGTGCTTAGGGGAATGCTGCTCCACCTTTACGGTGACCCCCAGGT CTGTGCTGAGCAGAACGCAGCTCCGCCCTCGCGGTACCCTCAGCCCGCCCGCCCGGGTCTGACCTGA GGAGAATTCTGCTCCGCCTTCGCAGTACCACCGAAATCTGTGCA	450
16p	AACCGACCCTCACCCTCACCCTCACCACATGAGCAATGTGGGTGTTATATTTTAGGTGTCATGTGTGC ATTAGGAATGCTGCATTTGTATTCCAACGCTGCAACTGGACCCTGCAATGCAGCCCCTCGCCTTGCCT TGGGAGAATCTCGGTGCGCAGGATTCAGAGGGGCTTTTAGTTTCCCGTTTTCCACACTGAACCGTTCT AACTGGTCTCTGACCTTGATTATTCACGGCTGCAACCGGGAAAGATTTTATTCACCGTCAATGCGCCC CGAGTTGTCCCAAAGCCAGGCAGTGCCCCCAACGTCTGTGCTTAGGAGAATGCTGCTCCACCTTTAC GGTGACCCCCAGGTCTGTGCTGAGCAGAACGCAGCTCCGCCCTCGCAGTAC	390
20q	CCCTAACCCTGATTATTCAGGGCTGCAAACAGAAAGGATTTTGTTCACTGTCGATGCGGCCCCGAGTT GTCCCAAAGCGAGGCTGTGCCCCCAAGGTGTGTGCTGAGAAAGCTGCTCCACCTTTGCGGTGTGCCC CGGGTCTGTGTTAAGCAGAACGCAGCTCTGCCCCCGCAAAGGCATGGCGCAGAGACGGGAGGAAC CTCAATAATTCGAAAAGCCGGGCTCGACCGCCCCCTGCTTGCAGCTGGGCACTACAGGACTGGCTTG GTCACGGTGCTGTGCCAGTGCACCCCCTGCTGGCGACTAGGGCAGCTGCAGGGCTCTGTTGCTTATA GTGGTGGCCAGCGCCCCCTGCTGGCGCCGGGGCAATGCAGGGGCCTCTTGCTCACAGTATA	395

Table A 2: sgRNAs designed for insertion in CRISPRi plasmid

Oligo	Sequence (5’ -> 3´)
7q-CpGisland-CRISPRi_Fw	caccGGGGCGCAGAAAAAGATCTA
7q-CpGisland-CRISPRi_Rev	aaacTAGATCTTTTTCTGCGCCCC
12q-CpGisland-CRISPRi_Fw	caccCGGGGCGCAGAGAAGATCTA
12q-CpGisland-CRISPRi_Rev	aaacTAGATCTTCTCTGCGCCCCG

Table A 3: primers used for RT-qPCR on chromosome ends XqYq and XpYp

Primer	Sequence (5’ -> 3´)	Ref.
XpYp Fw	CCCATTGGGTTGCCACTGCTGGCT	[105]
XpYp Rev	CCCTCTGAAAGTGGACCTAT	
XqYq Fw	TAAAAATGTTTCCCGGTTGC	[77]
XqYq Rev	CACCCTCACCCTAAGCACAT	

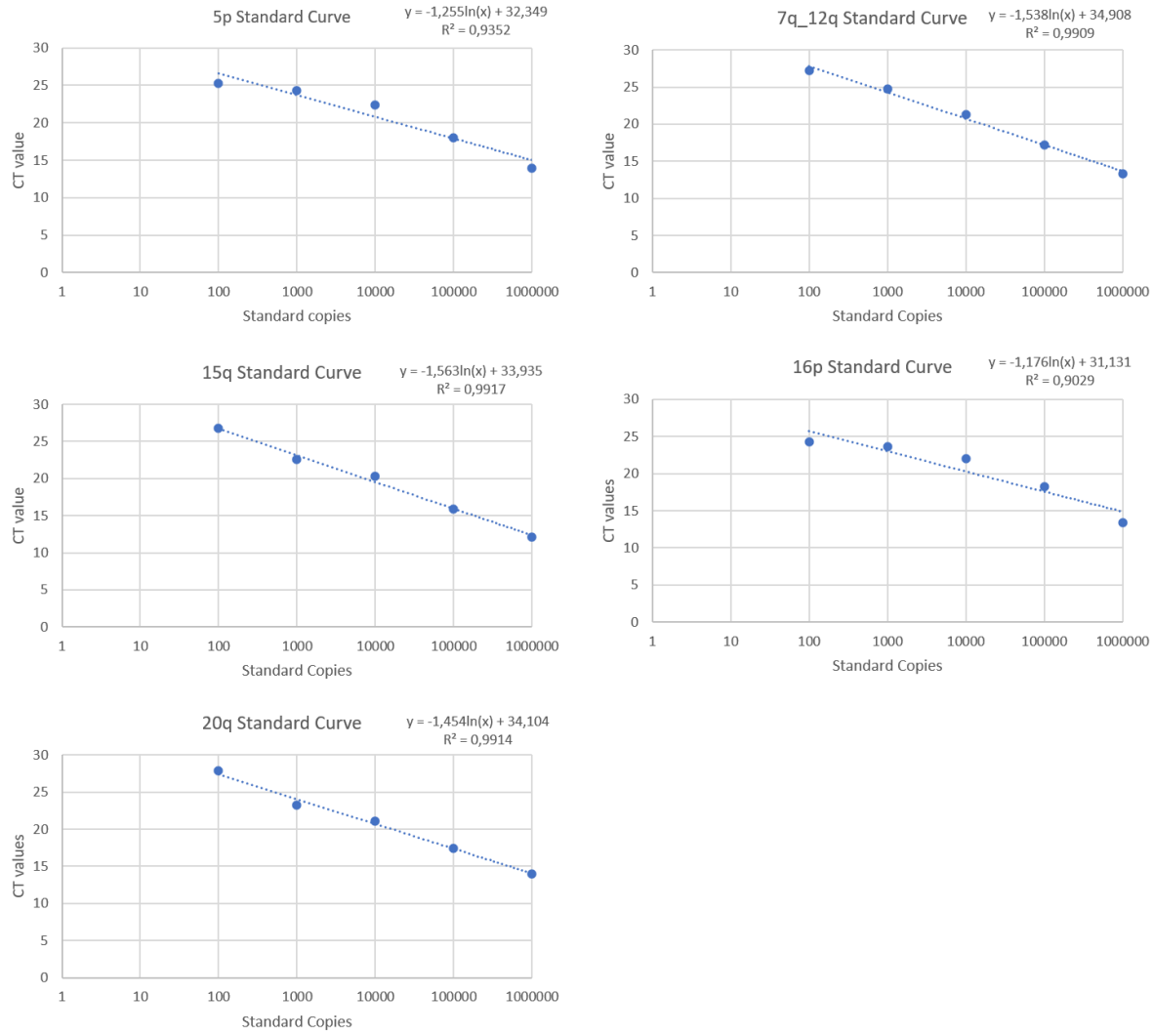

Figure A 1: Standard Curves of each of the tested chromosome ends, 5p, 7q_12q, 15q, 16p and 20q. y axis contains the CT values and x axis contains the number of standard copies in the reaction.

Table A 4: Motifs of transcriptional regulatory elements found in the 5000 bp before the first telomeric repeat of the 7q and 12q subtelomeres

Number	TF	Chromosome Ends	
		7q	12q
1	FOXD1	1	2
2	FOXC1	14	15
3	FOXL1	6	3
4	FOXI1	0	1
5	MZF1_1-4	9	16
6	MZF1_5-1	1	1
7	SOX9	1	1
8	SPIB	5	8
9	SRY	3	4

10	REL	1	1
11	RELA	1	1
12	NKX3-1	2	2
13	ZNF354C	9	9
14	BRCA1	5	6
15	NFIC	5	6
16	BATF::JUN	0	1
17	CEBPB	1	2
18	DUX4	0	1
19	E2F6	1	1
20	FOSL2	0	1
21	ELF	1	0
22	FLI1	1	0
23	FOXH1	3	4
24	MAFK	1	1
25	MEF2C	1	0
26	NRF1	1	2
27	POU2F2	1	0
28	RUNX2	1	1
29	TCF7L2	1	1
30	USF2	1	2
31	CEBPA	3	3
32	EBF1	3	3
33	EGR1	1	1
34	FOXA1	2	1
35	GATA3	1	2
36	MAX	3	1
37	MEF2A	1	0
38	NFKB1	2	3
39	SP1	4	5
40	STAT1	0	1
41	STAT3	3	3
42	TAL1::GATA1	0	1
43	USF1	1	1
44	YY1	1	0
45	ZEB1	3	3
46	FOXP2	1	2
47	EHF	0	1
48	KLF5	4	6