Biological Functions of G-Quadruplexes

Telomere Maintenance and Transcriptional Regulation in Embryonic Stem Cells

Rikard Runnberg

Department of Medical Biochemistry and Cell Biology
Institute of Biomedicine
Sahlgrenska Academy at the University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2014

Biological Functions of G-Quadruplexes
© Rikard Runnberg 2014
rikard.runnberg@gu.se

ISBN 978-91-628-9243-2
ISBN 978-91-628-9244-9 (*Electronic version*)
http://hdl.handle.net/2077/37104

Printed in Gothenburg, Sweden 2014
Aidla Trading AB/Kompendiet

"Character cannot be developed in ease and quiet. Only through experience of trial and suffering can the soul be strengthened, vision cleared, ambition inspired, and success achieved."

Helen Keller

Biological Functions of G-Quadruplexes

Telomere Maintenance and Transcriptional Regulation in Embryonic Stem Cells

Rikard Runnberg

Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine
Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden

ABSTRACT

G-quadruplexes are four-stranded nucleic acid structures formed by quartets of Hoogsteen base paired guanine bases. They are known to form at telomeres and other genomic sites, and are predicted to do so at a large proportion of gene promoters. They may function in telomere maintenance and transcriptional regulation, but their biological functions remain largely unknown. The aim of this thesis was to study the association of proteins with telomeres and telomeric G-quadruplexes, and to study protein-protein interactions in embryonic stem cells (ESCs).

This thesis contains three papers. In the first paper the association of heterogenous ribonucleoprotein U/Scaffold Attachment Factor A (hnRNP U/SAF-A) with telomeres was identified by *in situ* Proximity Ligation Assay (PLA) and chromatin immunoprecipitation (ChIP). It was found that hnRNP U associates with telomeres in a cell cycle dependent manner. DNA pull-down and exonuclease protection assay showed that hnRNP U via its C-terminus binds and promotes the formation of telomeric DNA G-quadruplexes, and that in doing so it can prevent RPA in ESC extracts from binding telomeric single-stranded DNA. Immunofluorescence (IF) following shRNA mediated knock-down of *Hnrnpu*, showed that hnRNP U also has a role in preventing RPA association with telomeres in cells. In the second paper IF, PLA and co-immunoprecipitation (co-IP) were used to identify hnRNP U and BRG1 as interaction partners in ESCs. Using an ethynyl uridine incorporation assay it was shown that both components are important for global transcription by RNA polymerase II. In the third paper protein affinity purification, IF, PLA and co-IP were used to identify interactions between nucleolin (Ncl) and two proteins in ESCs. Phosphorylated Ncl interacts with Tpt1 during mitosis and with Oct4 during interphase.

In this thesis hnRNP U is identified as a novel telomere binding protein. The results presented here suggest G-quadruplex formation may be an important aspect of telomere maintenance. Three novel protein-protein interactions were identified in ESCs. The identified protein complexes may have roles in key aspects of ESC biology, such as transcription and cell cycle regulation.

Keywords: G-quadruplex, telomere, embryonic stem cells, transcription, hnRNP U, BRG1, Ncl, Tpt1, Oct4

ISBN: 978-91-628-9243-2

ISBN: 978-91-628-9244-9 (*Electronic version*), *http://hdl.handle.net/2077/37104*

Sammanfattning på svenska

Arvsmassa lagras i molekylen deoxiribonukleinsyra (DNA). Det är en lång molekyl med en ryggrad bestående av sockerarten deoxiribos sammanlänkad med fosfodiesterbindningar. Varje deoxiribos bär en nukleinsyrabas. Vanligtvis bildar DNA molekylen en dubbelsträngad helix (d.v.s. en spiralform) i vilken baserna binder de två DNA-strängarna samman. Endast komplementära baser kan binda varandra i denna struktur. Baserna i DNA och deras ordning utgör den genetiska kod som styr cellens funktioner. Gener transkriberas till en snarlik enkelsträngad molekyl; ribonukleinsyra (RNA). RNA translateras i sin tur till protein som sedan utför funktioner i cellen. Då i stort sett alla celler i människan bär på samma gener måste olika celltypers funktion styras genom att olika gener uttrycks olika mycket. Vad som styr geners uttryck, samt hur celler kan skydda sitt DNA, är grundläggande frågor inom molekylär cellbiologi. I denna avhandling diskuteras processer som styr detta, samt hur den alternativa fyrsträngade DNA-strukturen G-quadruplex är involverad i dessa.

Det första delarbetet handlar om telomerer, d.v.s. ändarna på DNA-molekylerna. I eukaryoter är DNA ordnat på linjära molekyler som med hjälp av proteiner packats till kromatin. Linjärt DNA medför två problem:

1) Replikationsmaskineriet som kopierar DNA inför celldelning kan inte kopiera ändarna på kromosomerna.
2) Cellen måste på något sätt kunna skilja ändarna från dubbelsrängsbott inne i DNA-molekylen.

Det första problemet medför att kromosomerna blir kortare och kortare när cellen delar på sig. Vid en viss gräns leder det till att de slutar dela på sig och hamnar i ett vilande tillstånd som kallas senescens. Detta händer med våra celler när vi blir äldre och tros vara en förklaring till varför vi åldras. Vissa celler kan dock fortsätta dela sig om och om igen. Detta sker genom att cellen uttrycker telomeras som förlänger telomererna. Stamceller har naturligt denna förmåga. De finns i olika organ och kan förse dessa med nya celler. Kapaciteten är dock begränsad varför vi ändå åldras. Stamceller kan också isoleras från ett tidigt stadium av utvecklingen och kallas då embryonala stamceller. Dessa har den unika förmågan att förutom att kunna dela sig oändligt många gånger också kunna bilda alla celltyper i den vuxna kroppen. Cancerceller har återfått förmågan att kunna dela sig oändligt många gånger, och kan göra detta ohämmat, vilket leder till bildandet av tumörer. Att de flesta celler i kroppen inte uttrycker telomeras kan därför ses som en mekanism för att hämma uppkomsten av tumörer. Förståelsen för hur

förlängning av telomerer regleras är därför viktig för att förstå hur både stamceller och cancerceller fungerar.

Dubbelstängsbrott på DNA kan uppkomma spontant eller efter yttre påverkan (t.ex. strålning). Cellen har ett system för att känna igen dessa och ifall de inte repareras slutar cellen att dela sig eller genomgår programmerad celldöd. Om telomerer känns igen som dubbelsträngsbrott kan det leda till att cellen försöker reparera dem, vilket medför att telomerer sätts ihop med varandra, eller att delar av en telomer överförs till en annan. Detta ger genetisk instabilitet, som kan bidra till utvecklingen av cancer. Dessutom kan det precis som ett vanligt dubbelsträngsbrott leda till att cellen slutar dela sig eller dör. Därför måste det finnas en mekanism i cellen för att skilja mellan telomerer och dubbelsträngsbrott.

I det första delarbetet i denna avhandling identifieras ett protein, hnRNP U/SAF-A, som binder till telomerer främst under början av den del av cellcykeln då DNA replikeras (d.v.s. kopieras). hnRNP U binder och stabiliserar G-quadruplex. Detta kan skydda telomerer från att kännas igen av proteinkomplexet RPA, som kan aktivera DNA-skadesignalering. Att G-quadruplex kan bildas vid telomerer som en biprodukt vid replikation och då få ödesdigra konsekvenser är välkänt. Men varför skulle telomerer ha en DNA-sekvens som (till skillnad från de flesta andra) kan bilda G-quadruplex om det är skadligt för cellen? Vårt arbete tyder på att G-quadruplex också kan ha en skyddande roll vid telomerer.

Det andra delarbetet handlar om hur hnRNP U kan interagera med proteinet BRG1, som har förmågan att förändra strukturen hos kromatin med hjälp av kemisk energi. Vi visar att båda dessa proteiner har en viktig roll för att transkription över huvud taget ska kunna ske i embryonala stamceller.

I det tredje delarbetet identifieras två proteinkomplex i embryonala stamceller som innehåller proteinet nucleolin (Ncl) och proteinerna Tpt1 respektive Oct4. Exakt vilken funktion dessa komplex har är fortfarande okänt, men möjligtvis reglerar de cellcykel och transkription. Ncl vet man sedan tidigare binder G-quadruplex, varför det kan tänkas ha betydelse även för dessa proteinkomplex.

Vi har visat att hnRNP U har en roll vid telomerer kopplad till bindning av G-quadruplex, och vi har identifierat tidigare okända protein-protein komplex. De resultat som presenteras här tyder på att G-quadruplex och proteiner som binder dessa har viktiga funktioner i embryonala stamceller.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. **Runnberg, R**., Vizlin-Hodzic, D., Green, L.C., Funa, K., Simonsson, T. hnRNP U/SAF-A is a G-quadruplex binding protein that associates with telomeres in a cell cycle dependent manner.
Submitted manuscript, under revision.

II. Vizlin-Hodzic, D., **Runnberg, R.**, Ryme, J., Simonsson, S., Simonsson, T. SAF-A forms a complex with BRG1 and both components are required for RNA polymerase II mediated transcription.
PLoS ONE 6(12): e28049.
doi:10.1371/journal.pone.0028049.

III. Johansson, H., Svensson, F., **Runnberg, R**., Simonsson, T., Simonsson, S. Phosphorylated nucleolin interacts with translationally controlled tumor protein during mitosis and with Oct4 during interphase in ES cells.
PLoS ONE 5(10): e13678.
doi:10.1371/journal.pone.0013678.

CONTENT

ABBREVIATIONS XI
DEFINITIONS IN SHORT XIV
1 INTRODUCTION 1
1.1 The G-quadruplex: a structure in search of function 1
1.1.1 History of the G-quadruplex 2
1.1.2 Structural diversity of G-quadruplexes 3
1.1.3 Small molecule G-quadruplex ligands 4
1.2 Embryonic Stem Cells 4
1.2.1 The ES cell cycle 5
1.3 Telomere biology 6
1.3.1 The end-protection problem 6
1.3.2 The end-replication problem 11
1.3.3 Telomere length regulation: longevity versus tumor suppression 12
1.3.4 Telomere maintenance as a whole 25
1.4 Transcriptional regulation 26
1.4.1 Chromatin and its remodeling 27
1.4.2 The role of G-quadruplexes in transcriptional regulation 29
2 AIMS 32
3 ASPECTS OF METHODOLOGY 33
3.1 Studying protein-DNA interactions 33
3.1.1 DNA pull-down 33
3.1.2 Chromatin immunoprecipitation 34
3.1.3 Immunofluorescence and fluorescence *in situ* hybridization 34
3.2 Studying protein-protein interactions 35
3.2.1 Protein affinity purification 36
3.2.2 Co-immunoprecipitation 36
3.2.3 Immunofluorescence 36
3.2.4 *In situ* proximity ligation assay 37

3.3 Assessing G-quadruplex stabilization 37
3.3.1 Exonuclease I protection assay 38
3.4 RNAi mediated gene knock-down 38
3.5 Studying transcription 39
3.5.1 Global transcription assay 39
3.5.2 Reverse transcription quantitative PCR 39
4 RESULTS 41
4.1 Paper I 41
4.2 Paper II 43
4.3 Paper III 44
5 DISCUSSION 45
6 CONCLUSION 47
7 FUTURE PERSPECTIVES 48
ACKNOWLEDGEMENT 50
REFERENCES 52

ABBREVIATIONS

ALT	Alternative lengthening of telomeres
ATM	Ataxia-telangiectasia mutated
ATR	Ataxia-telangiectasia and Rad3-related
ATRIP	ATR interacting protein
BRG1	Brahma related gene 1
ChIP	Chromatin immunoprecipitation
Co-IP	Co-immunoprecipitation
CST	Ctc1-Stn1-Ten1
CTD	C-terminal domain
DAPI	4,6-diamino-2-phenylindole
DDR	DNA damage response
DMS	Dimethyl sulfate
DNA	Deoxyribonucleic acid
cDNA	complementary DNA
dsDNA	Double-stranded DNA
rDNA	Ribosomal DNA
ssDNA	Single-stranded DNA
DKC1	Dyskerin
EMSA	Electrophoretic mobility shift assay
ESC	Embryonic stem cell
EU	5-ethynyl uridine
FISH	Fluorescence *in situ* hybridization
FRET	Förster resonance energy transfer
FSHD	Facioscapulohumeral muscular dystrophy
hnRNP U	Heterogenous ribonucleoprotein U
HAT	Histone acetyltransferase
HDAC	Histone deacetylase

HMT	Histone methyltransferase
HP1	Heterochromatin protein 1
ICM	Inner cell mass
IF	Immunofluorescence
iPSC	Induced pluripotent stem cell
kb	Kilobase (i.e. 1000 bases)
MEF	Mouse embryonic fibroblast
MRN	MRE11-RAD50-NSB1
Ncl	Nucleolin
NHE	Nuclease hypersensitive element
NHEJ	Non-homologous end joining
Npm1	Nucleophosmin
Oct4	Octamer binding transcription factor 4
PLA	*In situ* proximity ligation assay
Pol	RNA polymerase
POT1	Protection of telomeres 1
PQS	Putative quadruplex forming sequence
RAP1	Repressor activator protein 1
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
mRNA	Messenger RNA
miRNA	Micro RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
RPA	Replication protein A
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SAF-A	Scaffold attachment factor A
S/MAR	Scaffold/matrix attachment region
TCAB1	Telomerase Cajal body protein 1

TERRA	Telomere repeat-containing RNA
TERC	Telomerase RNA component
TERT	Telomerase reverse transcriptase
TIF	Telomere dysfunction induced foci
TIN2	TRF1-interacting nuclear protein 2
TMPyP4	G-quadruplex ligand 5,10,15,20-Tetrakis(1-methylpyridinium-4-yl)porphyrin tetra(p-toluenesulfonate)
TPE	Telomere positioning effect
TPP1	TIN2 and POT1-interacting protein
Tpt1	Tumor protein translationally controlled 1
TRAP	Telomere repeat amplification protocol
TRF1	Telomere repeat binding factor 1
TRF2	Telomere repeat binding factor 2
T-SCE	Telomere-sister chromatid exchange
TSS	Transcription start site
53BP1	p53 binding protein 1

DEFINITIONS IN SHORT

Term	Definition
Embryonic stem cell	A pluripotent stem cell isolated from the inner cell mass of an embryo.
Exon	Part of a gene that remains after RNA splicing.
G-quadruplex	A four-stranded nucleic acid structure where planes of four guanine bases are Hoogsteen base paired. Formed by repeats of G-bases.
Helicase	Enzyme that separates the strands of DNA, RNA or DNA-RNA by ATP-hydrolysis.
Heterochromatin	Densely packed chromatin.
Holoenzyme	An enzyme containing many different subunits.
Interphase	The part of the cell cycle when the cell is not undergoing mitosis.
Ligand	A substance that forms a complex with a biomolecule.
Mitosis	The part of the cell cycle where a cell divides.
Pluripotency	The ability to form all cell types (germ layers) of the body.
Proto-oncogene	A gene which upon its activation induces cancer.
Replication	The process by which DNA is copied.
Senescence (cellular or replicative)	The process by which a cell stops dividing after a certain number of divisions.

Somatic cell	All cells of the body except germ cells.
Stem cell	A cell that can to divide an infinite number of times and differentiate into other cell types.
Supercoiling	The over- or under-winding of DNA.
Telomerase	The enzyme that lengthens telomeres.
Telomere	The nucleoprotein structure at the end of a chromosome.
Transcription	The process by which DNA is copied to RNA.

1 INTRODUCTION

1.1 The G-quadruplex: a structure in search of function

Deoxyribonucleic acid (DNA) is the biomolecule that carries genetic information. Eukaryotic DNA is localized in the nucleus of cells on linear chromosomes. Genes are transcribed from DNA into ribonucleic acid (RNA), which is transported out of the nucleus to be translated into proteins that carry out most of the biological functions inside and outside of cells. The bases of DNA make up codons for specific amino acids, the building blocks of proteins. The knowledge that DNA may contain the blueprint for all biological processes led to a lot of interest in its structure. It was first solved through cleaver model building by James Watson and Francis Crick (1). The canonical structure of DNA is the Watson-Crick base paired B-form double helix. This DNA double helix has become somewhat of an icon for molecular biology. The major importance of the molecule is the code it contains. Thus, one might think that now that we also know the entire sequence of many genomes including the human, all the answers would be in this code. However, since all cells of an organism contain the same genetic information in its DNA (with few exceptions), diversity is created by what genes are being expressed. How this is controlled is one of the main questions in biology today. Another important aspect of DNA is how it can be copied without errors and without accumulating devastating damage that may for instance lead to cancer. One of the most devastating types of damage is double-strand breaks. How can the ends of a linear chromosome be distinguished from such breaks?

DNA is known to form several different non-canonical structures, of which the biological function is not as well known, but that may help control gene expression and maintain the integrity of the DNA. These include the alternative A- and Z-form double helices as well as three and four-stranded structures. The most well studied four-stranded structure, and the one that is likely to have the greatest physiological relevance, is the G-quadruplex. It has enticed structural biologists for over a century, but finding its possible biological functions has remained elusive. Only recently have G-quadruplexes been firmly proven to even exist in mammalian cells, and now evidence is building up showing they have fundamental roles in cell biology.

1.1.1 History of the G-quadruplex

Already in 1910 it was reported that concentrated solutions of guanylic acid forms a gel (2). During the 1960s, when X-ray diffraction was established as a tool for studying the structure of biomolecules, this gel was found to consist of planar tetramers of guanines connected by Hoogsteen base pairing (3). This has been termed a G-quartet or G-tetrad (figure 1) (4,5). Over a decade later it was found that repeats of guanines in polyguanylic acid (poly(G)) and polydeoxyguanylic acid (poly(dG)) can form four-stranded structures called G-quadruplexes, due to the formation of such stacked G-quartets (6-8). The first endogenous sequences found to form G-qudruplexes in biochemical experiments were immunoglobulin switch regions (9). Walter Gilbert and Dipankar Sen proposed G-quadruplexes could aid the alignment and recombination of chromosomes during meiosis and that they may also form at telomeres, which were known to be G-rich in most species studied and to exist as single-stranded DNA in 3' overhangs (9). In a study from Elizabeth Blackburn's lab, a double G-G base paired hairpin was proposed to be formed by *Tetrahymena* telomeric DNA (10), but the year after Sen's and Gilbert's suggestion was published, studies from the labs of Thomas Cech and Aaron Klug showed that the telomeric sequence from the ciliates *Oxytricha* and *Tetrahymena* do indeed form G-quadruplexes (11,12). Subsequently a G- quadruplex was found to form also in a control region of the proto-oncogene c-Myc (13), suggesting a function in transcriptional regulation. Further indication of a role for G-quadruplexes in transcription came when bioinformatics were applied to look for putative quadruplex forming sequences (PQSs) in different genomes. It was found that their distribution is not random. They are largely absent in exons, but are frequently found upstream of transcription start sites (TSSs) (14,15). From then on the G-quadruplex field has continued to grow. The study of their biological roles has however been hampered by the difficulties in assessing G-quadruplex function *in vivo*, and to actually visualize them *in situ*. Only recently were antibodies used to visualize G-quadruplexes throughout the mammalian genome (16,17). There is

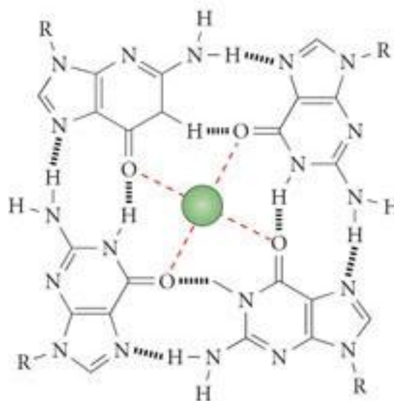

Figure 1. G-tetrad stabilized by a monovalent cation (green). Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Drug Discovery (Balasubramanian, S. *et al*, "Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?", vol. 10, p. 261-275), copyright (2011).

convincing indirect evidence of the importance of G-quadruplexes *in vivo*, acquired through characterization of proteins that resolve or stabilize G-quadruplexes and their functions in living cells. Those biological functions will be further discussed in subsequent chapters of this introduction.

1.1.2 Structural diversity of G-quadruplexes

The prerequisite for a DNA sequence forming a G-quadruplex is that it contains repeats of G-bases. It can form inter- or intra-molecularly, where formation of the latter requires four repeats of G-bases. G-quadruplexes are stabilized by monovalent cations positioned in the center of the G-tetrads, coordinating the oxygens pointing inwards (figure 1). The stabilizing effect of cations seems to rely on their diameter, where K^+ and Na^+ are the most efficient stabilizers (that are also physiologically relevant), while the much smaller Li^+ inhibits formation of G-quadruplexes (18,19). The orientation of the strands in a quadruplex can vary. The strand polarity dictates whether the glycosilic bonds in each strand will be in the *anti* or *syn* conformation, which affects the size of the G-quadruplex grooves (figure 2) (5). In any G-quadruplex made up of less than four DNA molecules, the tetrads will be

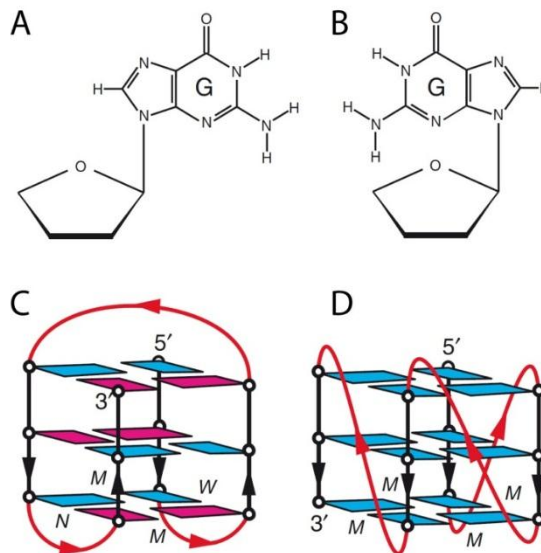

Figure 2. (A) Anti and (B) syn conformation of the glycosylic bond in guanosine (the OH-groups have been left out). (C) Scematic figure showing the "basket" telomeric G-quadruplex observed in Na^+ solution. (D) Schematic figure showing the "propeller" quadruplex observed in a K^+ containing crystal. N – narrow, M – medium, and W-wide groves. Adapted from Phan, A. T. (2010), FEBS Journal, "Human telomeric G-quadruplex: Structures of DNA and RNA sequences", 277: 1107-1117. doi: 10.1111/j.1742-4658.2009.07464.x, © 2009 The Author Journal compilation © 2009 FEBS , with permission from John Wiley and sons.

connected by loops, which can run on top or on the outside of the quadruplex. Thus, there is plenty of structural variation to G-quadruplexes. Structure is sequence dependent, but some sequences also fold in more than one way. This is the case for the human telomeric sequence where many different inter- and intramolecular structures have been observed, including a K^+-stabilized parallel, “propeller”, and a Na^+-stabilized antiparallel, “basket”, G-quadruplex (figure 2) (20).

1.1.3 Small molecule G-quadruplex ligands

Part of the interest in understanding the structures and biological functions of G-quadruplexes lies in the potential of interfering with their functions through the use of small molecule ligands. A multitude of such ligands have been developed. One of the most well characterized is the cationic porphyrin 5, 10, 15 , 20-Tetrakis(1-methylpyridinium-4-yl)porphyrin tetra(p-toluene-sulfonate) (TMPyP4) developed as a ligand for the human telomeric G-quadruplex (21). It was designed to stack on the outer G-tetrad of the quadruplex, but it has later been shown to instead stack on the TTA nucleotides (22). Using heteroaromatic molecules that fit on top of the G-tetrad is a common theme among G-quadruplex ligands to minimize interaction with dsDNA and ssDNA in its extended conformation (4). Specificity for a certain quadruplex might be increased by side chains that fit between the loops and grooves of the quadruplex (4). A review of all G-quadruplex ligands is beyond the scope of this thesis, but suffice to say that while no G-quadruplex ligand has been approved for clinical use yet, they show great potential as cancer therapeutics and have already found widespread use in G-quadruplex research.

1.2 Embryonic Stem Cells

All the papers presented in this thesis study murine embryonic stem cells (mESCs). These cells carry a number of unique characteristics, which have spurred a lot of interest in understanding their fundamental cell biology.

A defining feature of stem cells is their ability to give rise to more differentiated cells. Pluripotent cells can form all germ layers (except for the extraembryonic trophectoderm), i.e. all cell types of the adult body, while multipotent cells can form more than one, and unipotent only give rise to one cell type. ESCs are pluripotent. These are cells isolated from the inner cell mass (ICM) of an embryo at the blastocyst stage (23). Their ability to grow indefinitely in culture and form any cell type could allow for potential applications in transplantation therapies, as well as in advanced *in vitro*

culture systems for studying disease. Protocols have been developed for inducing pluripotency in differentiated cells (24,25), creating induced pluripotent stem cells (iPSCs), having all the characteristics of embryonic stem cells, but enabling personalized therapies and disease specific model systems, while avoiding the ethical controversy of using human embryos. ESCs have a unique cell cycle reflecting the demands of rapid proliferation needed in the early stages of development from which they were isolated. Also, they have the unique property of being able to self-renew, i.e. to divide an infinite number of times, which is further discussed in the chapter "The telomere end-replication problem".

1.2.1 The ES cell cycle

The normal cell cycle consists of mitosis, the division of the cell nucleus, and the time in between that is called interphase. The cell cycle is divided into four phases in the order G1 – the first gap phase, S – the phase where DNA replication occurs, G2 – the second gap phase, and M – mitosis. During mitosis the tetraploid genome that exists following S-phase needs to be accurately divided into two cells each containing a full set of the diploid genome. This occurs in five phases: prophase – chromosomes containing sister chromatids start condensing and the mitotic spindle starts forming, prometaphase – the nuclear envelope breaks down and microtubuli start attaching to each chromosome, metaphase – chromosomes align, anaphase – sister chromatids are pulled apart, and telophase – daughter chromosomes reach each pole of the mitotic spindle, start decondensing and new nuclear envelopes are built. This is followed by cytokinesis where the cytoplasm is divided in two. Then G1 begins again in the two cells.

Transition between cell cycle stages is a regulated process ensuring that cell cycle events occur in the right order. This regulation also ensures that there are checkpoints, so that the cell cycle can be stopped upon DNA-damage, as will be discussed later. The main proteins involved in regulating the cell cycle are a set of proteins called cyclin dependent protein kinases (Cdks). These form complexes with regulatory subunits called cyclins. Cdk-cyclin complexes affect downstream targets to drive the cell cycle, and the activity of each cell cycle phase specific Cdk-cyclin is regulated in a precise manner (26).

ES cells are unique in that they have a truncated G1-phase (27). In somatic cells Cdk2 activity regulates the G1/S transition and progression through S-phase. mESCs have constitutive Cdk2 activity, owing to constantly high levels of cyclin A and E, and low levels of Cdk-inhibitory proteins p21 and

p27 (28). Retinoblastoma protein (Rb), a target of Cdk2-cyclin E, is constantly phosphorylated and thus incapable of inhibiting G1/S transition (27). Cdc2-cyclin B on the other hand shows the normal pattern of high activity only during G2/M (28). This truncation of G1 leads to a considerably shorter cell cycle, but also changes in the checkpoints activated by DNA-damage, as discussed below.

1.3 Telomere biology

Telomeres are the nucleoprotein complexes that cap the ends of eukaryotic chromosomes. The DNA sequence of telomeres is conserved in most eukaryotes and consists of a G-rich strand and a complimentary C-rich strand (29). In vertebrates the conserved sequence is repeats of TTAGGG (30). The total length of telomeres varies between species and between different cell types in the same species. Human telomeres are typically between a few and up to 14 kilobases (kb) in length (31), while in laboratory mice they are up to 150 kb long (32). Telomeres typically contain a 3’ overhang on the G-rich strand that ranges between 50 and 300 bases in mammalian cells (33). The evolution of the eukaryotic genome to encompass linear rather than circular chromosomes has introduced a number of challenges on the cellular level, of which I will give a brief overview in the following chapters.

1.3.1 The end-protection problem

A challenge of linear eukaryotic chromosomes is that their ends must not be recognized and processed as DNA breaks. DNA damage occurs spontaneously in cells, and intricate machineries are in place for dealing with it, collectively termed the DNA damage response (DDR). It can result either in growth arrest, cell death, or repair. DNA damage repair at telomeres can lead to fusion of chromosomes and further genomic instability, which is an important factor in the development of cancer. The DNA-damage responses that need to be prevented at telomeres have been termed the end-protection problem.

The DNA damage response at telomeres

DNA damage signaling starts with the detection of the DNA damage by a set of proteins, followed by downstream signaling that result in repair, cell cycle arrest, or cell death. The two main pathways acting at DNA double-strand breaks are those mediated by ataxia- telangiectasia mutated (ATM) and ataxia-telangiectasia and Rad3-related (ATR) kinases. At telomeres these can act independently from one another, presumably due to the presence of the long 3’ overhang (34).

Unprotected telomeres can be recognized by the MRE11-RAD50-NBS1 (MRN) complex which activates ATM kinase (35,36). Single-stranded DNA and thus the telomeric overhang can be bound by replication protein A (RPA), which can recruit ATR via ATR Interacting protein (ATRIP) (37). At the sites of DNA damages, telomere dysfunction induced foci (TIF), containing phosphorylated histone H2AX (γ-H2AX) and p53 binding protein 1 (53BP1), among others, are formed (35). ATR activation leads to a signaling cascade that involves the phosphorylation of both Chk1 and Chk2 kinases, while ATM signaling leads primarily to Chk2 phosphorylation (34). Chk1 and Chk2 then phosphorylate downstream targets such as Cdc25 phosphatase (38), which causes cell cycle arrest, and transcriptional regulator p53, which causes apoptosis or cell cycle arrest (36,39). Repair of the DNA damage can proceed either via homologous recombination or non-homologous end joining (NHEJ).

Less is known about the DDR in mESCs, especially in the context of telomere protection, but it is known to differ from that of somatic cells. Early responses to DNA damage are functional, but some of the downstream signals are not, resulting in a high degree of apoptosis instead of G1/S arrest (40). Instead of functioning in the nucleus upon DNA damage, Chk2 is retained at centrosomes (41). Chk1 on the other hand is functional and essential for a G2 arrest (42). p53 was first reported to not function in mESCs (40), but more recent reports suggest that it can regulate apoptosis and differentiation (43,44). High sensitivity to DNA-damage and increased apoptosis might lead to the very low rate of mutations observed in mESCs (45). Thus, mESCs are expected to also be sensitive to DDR originating from telomeres. In agreement with this, MEFs deficient in some of the known telomere protecting factors (discussed further below) are inefficiently reprogrammed to iPSCs (46,47).

Proteins that protect telomeres

The end-protection problem is solved primarily by proteins that specifically bind to and protect telomeres, but also proteins of the general DNA damage repair machineries (48). The core telomere binding protein complex is the six-membered *Shelterin* complex (49). It comprises telomere repeat binding factors 1 and 2 (TRF1 and TRF2) that bind double-stranded telomeric DNA as homodimers (50,51). The 3' overhang is bound by protection of telomeres 1 (POT1, or POT1a and POT1b in mouse) (52). TRF1-interacting nuclear protein 2 (TIN2) (53) interacts with both TRF1 and TRF2, and together with POT1 and TIN2-interacting protein (TPP1, also known as ACD, PTOP, TINT1 and PIP1) tethers the ssDNA binding to the dsDNA binding part of

the complex (54). TRF2 is bound by repressor activator protein 1 (RAP1), the sixth member of the complex (55) (figure 3).

TRF2 is perhaps the most essential end-protecting protein as it is required both for preventing ATM signaling and NHEJ, and much of what is known about the DDR at telomeres comes from studying this protein (36,56). POT1, and specifically POT1a in mice, on the other hand inhibits ATR signaling by preventing RPA from binding the ssDNA of the telomeric overhang (34,57).

The most simplistic model for how the Shelterin complex protects telomeres is that it works as a protective cap to prevent recognition by proteins of the DDR. While this may be partly true, it is overly simplistic as evident by the fact that many of the factors in DDR are actually needed at telomeres.

ATM and the MRN complex associate with functional telomeres in a cell cycle dependent manner, coinciding with loss of Pot1 and increased accessibility of telomeres in G2 (58). Members of the ATR mediated DDR, such as RPA, are also found transiently at telomeres during late S-phase of the cell cycle (59). These factors have been proposed to be required for resolving secondary DNA structures during replication, and then allowing telomeres to be properly processed during G2 (58,59). This hypothesis was strengthened by the subsequent findings that helicases such as WRN, BLM, RTEL1, and RECQL4, found mutated in Werner, Bloom, Hoyeraal-Hreidarsson, and Rothmund-Thomson syndrome respectively, are all needed for maintaining telomere integrity (60-63). Thus, it is clear that the composition of telomeres must be dynamic, rather than relying on a fixed capping structure.

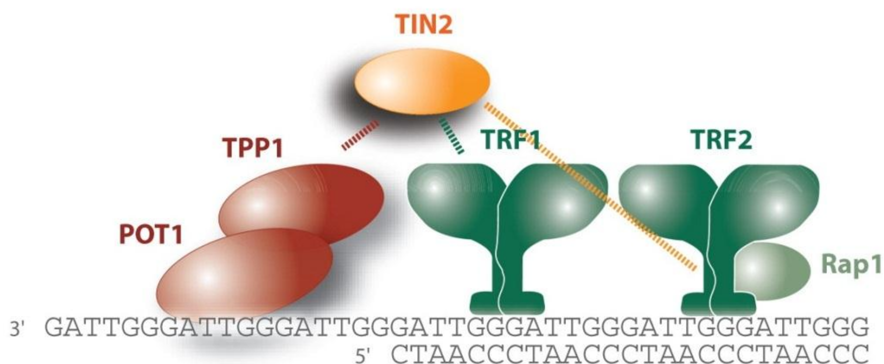

Figure 3. Schematic figure of the Shelterin complex bound to telomeric DNA. Reprinted from Palm, W. and de Lange, T. (2008), "How Shelterin Protects Mammalian Telomeres", Annual Review of Genetics, Vol. 42: 301-334, copyright 2008 Annual Reviews.

The association of RPA with telomeres highlights a gap in the understanding of telomere protection by Shelterin. If RPA needs to associate with telomeres, it also needs to be removed at a subsequent part of the cell cycle. Genetically depleting Pot1 from cells leads to accumulation of RPA at telomeres, but at physiological concentrations of RPA, which are very high, Pot1-Tpp1 cannot alone compete for binding to single-stranded telomere DNA *in vitro* (64). Hence, additional factors are required. This was shown to be facilitated in part by another abundant protein, hnRNP A1, which in its turn is removed by competition from telomere repeat-containing RNA (TERRA) (64). Hence, abundant proteins involved in general processes related to DNA and RNA, such as RPA and hnRNP A1, as well as telemetric transcripts, are required for solving the end-protection problem. The transcription of telomeres (to produce TERRA) is also a regulated process, both in development and during the cell cycle (65,66). This provides a clear link between the processes of transcription and replication in the protection of telomeres.

Alternative DNA structures that protect telomeres

The requirement of helicases specifically at replicating telomeres, suggest that the telomeric DNA sequence has a high tendency to form stable secondary structures. WRN, BLM, RECQL4 and RTEL1 can resolve two structures believed to form at telomeres; T-loops and G-quadruplexes. While it is clear that such structures need to be resolved, it seems paradoxical that the general features of telomeric DNA needed for forming these structures, i.e. being repetitive, G-rich, and containing a 3’ overhang, would be conserved among eukaryotes. It stands to reason that such structures must contribute to telomere maintenance.

T-loops are lariat-like structures where the G-rich overhang invades the double-stranded part of the telomere, thus forming a big loop referred to as the T-loop and a smaller loop where the complimentary G-rich strand is displaced, called a D-loop (figure 4). This has been proposed to hide the 3’ end of telomeres from the DDR machinery. TRF2 promotes the formation of T-loops, which might partially explain how it functions to protect telomeres (67-69). T-loops have been identified in a subset of telomeres on isolated chromosomes by both electron microscopy and super resolution fluorescence microscopy (68,69). The size of the T-loops observed by super resolution microscopy varied from about 7 to 30 kb, averaging 13 kb in mouse embryonic fibroblasts, and thus encompasses a large portion of the total telomere length (69). It remains unclear why T-loops are not present at every telomere. It was suggested to be a consequence of the isolation procedure, but it seems likely that the T-loop is in fact not the only protective secondary structure found at telomeres. RTEL1 knock-out leads to telomere loss due to

T-loop sized deletions, but it also leads to telomere fragility (62). Telomere fragility is defined by the detection of multiple telomere foci at one chromosome end in telomere-fluorescence *in situ* hybridization (FISH), that are believed to result from replication fork stalling (70). G-quadruplex stabilization either by knock-down of BLM or treatment with G-quadruplex ligand TMPyP4 exacerbated the telomere fragility phenotype of RTEL1 knock-out, but not the telomere deletion resulting from excised T-loops (62). Hence, there appears to be at least two separate secondary structures forming at telomeres; T-loops and G-quadruplexes, causing telomere deletion or fragility respectively, if not efficiently unwound during telomere replication. This shows that G-quadruplexes form during replication when the G-rich and C-rich strands are separated. It does however not exclude that they could also form during other parts of the cell cycle at the telomeric 3' overhang. Also, T-loops and G-quadruplexes are not mutually exclusive, as a G-quadruplex could form at the displaced strand of the D-loop, or at a part of the overhang that has not hybridized to the C-rich strand. The stability of such a G-quadruplex does not however appear to be crucial for the T-loop sized deletions in RTEL1 knock-outs.

Recently, G-quadruplexes have been observed at mitotic mammalian chromosomes, by two independent studies using two different G-quadruplex antibodies (16,17). The amount of G-quadruplex staining differed somewhat between the two studies. About 25% of mitotic chromosome ends were stained in the first study, while many chromosome ends were also stained in the second study (although no quantification was presented in the latter) (16,17). Roughly a third of all telomeres in interphase U2-OS cells colocalized with G-quadruplexes (16). Thus, G-quadruplexes at telomeres are unlikely to only exist at a transient state during replication. These were the first studies showing G-quadruplexes *in situ* using antibodies in mammalian cells.

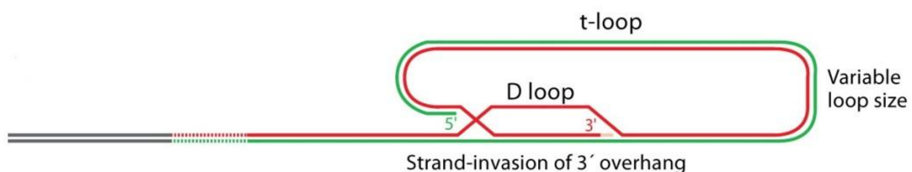

Figure 4. Schematic figure showing how a T-loop may form at the end of a telomere. Reprinted from Palm, W. and de Lange, T. (2008), "How Shelterin Protects Mammalian Telomeres", Annual Review of Genetics, Vol. 42: 301-334, copyright 2008 Annual Reviews.

G-quadruplexes have been suggested to protect telomeres by inhibiting RPA binding. RPA is very abundant in the cell, while Pot1 is not, but even when present in equimolar amounts POT1-TPP1 cannot compete efficiently for binding telomeric ssDNA (71). However, when a K^+-stabilized parallel G-quadruplex forms next to the ss/dsDNA border of a 3' overhang, with POT1-TPP1 bound at the distal end, RPA only outcompetes POT1-TPP1 at about haft of the overhangs (71). Both RPA and POT1-TPP1 bind telomeric ssDNA in its unfolded conformation through OB-fold domains (72,73), and both can unfold telomeric G-quadruplexes (74,75). However, stabilizing the G-quadruplex with small-molecule ligands inhibits the binding of both proteins (71,76). Given these similarities it seems counter intuitive that a G-quadruplex would aid in the binding of one over the other. This might be explained by their binding polarities and footprints. Pot1 binds telomeric ssDNA and unfolds G-quadruplexes in the 3' to 5' direction, while Pot1-Tpp1 slides in both directions (74). RPA consists of three subunits (RPA1, 2, 3). It binds and unfolds the G-quadruplex in a 5' to 3' direction, binding first with the RPA1 subunit at the 5' end, then with RPA2 3' of this binding site (77). RPA does however have different binding modes encompassing different domains and different sized footprints (78), and the unfolded ssDNA at the 3' end of a telomeric G-quadruplex greatly improves its ability to resolve the G-quadruplex (71). Notably, half of the overhangs were bound by RPA even when POT1-TPP1 could compete for binding, suggesting that additional factors are required *in vivo* (71). Such a factor could be a G-quadruplex stabilizing protein.

1.3.2 The end-replication problem

A problem appeared when the mechanisms of DNA-replication began to unravel in the early 1970's. DNA-polymerase requires an RNA primer to initiate replication, leaving the 5' end of the lagging strand shortened after every round of replication (79). This was termed the "end-replication problem". If left unsolved this problem would lead to genetic loss and potentially cell death as soon as the cell had divided a finite number of times. Intriguingly, this phenomenon termed cellular senescence had previously been observed in cell culture, where human cells would stop dividing after around 50 population doublings (80). On the organismal level there must be a solution to this, otherwise the long term survival of any eukaryotic species would be impossible. The solution was discovered on Christmas day in 1984 by Carol Greider, working in the lab of Elizabeth Blackburn, who had herself first discovered the repetitive sequence of telomeres about one decade earlier. In a cell extract from the ciliate Tetrahymena, she found an enzymatic activity that added telomeric repeats in a processive manner. The enzyme was

named telomerase, and for the discovery of telomeres and telomerase Greider, Blackburn and their collaborator Jack Szostak, were awarded the Nobel Prize in physiology or medicine in 2009.

Later, a mechanism for alternative lengthening of telomeres (ALT) was discovered that does not require telomerase, instead being facilitated by recombination (81). While ALT is rare compared to telomerase expression in tumors (81), a telomerase independent mechanism for telomere lengthening can also be activated in yeast that normally rely on telomerase (82). This seems to be at play also in early mammalian development (83), suggesting there are conserved back-up mechanisms for rapid telomere lengthening.

1.3.3 Telomere length regulation: longevity versus tumor suppression

As telomerase was further characterized it was found that mutations in yeast telomerase lead to telomere shortening and cellular senescence similar to that observed when culturing mammalian cells (84). It was also confirmed that the number of cell divisions before senescence in human fibroblasts indeed correlates with telomere length (85), while reconstitution of telomerase leads to elongation of telomeres and increased lifespan (86). The telomere induced senescence observed in human fibroblasts is similar to that seen for telomere dysfunction upon faltering end-protection and involves the DNA damage response and downstream signaling, leading to the activation of p53 and p21 (39,87). In mice it was found that primary cell cultures and most tissues lack telomerase activity, and telomere length in different tissues decreases with aging (88). A decrease in telomere length has also been observed in human tissues (89). This clearly indicates that cellular senescence observed in culture is relevant to organismal aging and that telomerase and telomere length are controlling factors.

Further evidence of this came with the discovery that mutations in telomerase components are found in patients suffering from the premature aging syndrome dyskeratosis congenita (90,91), which mainly affects tissues that require continuous renewal such as skin, gut and bone marrow. Symptoms include hyperpigmentation of the skin, dystrophy of nails, premature graying, etc. A role for telomerase in aging was confirmed when the first telomerase knock-out mice were created, and shortened telomeres, telomere fusions (92), shortened life spans (93), and symptoms similar to dyskeratiosis congenita were observed (94). These phenotypes manifest mainly after a few generations, suggesting that at least laboratory mice have a certain reserve telomere length (92-94).

It stands to reason that, at least during some point of development, telomeres must be lengthened. Otherwise the same phenotypes observed in telomerase knock-out mice would be expected. Early on it was found that the testes, i.e. male germ cells, have high telomerase activity (88). Later it was also found that telomeres lengthen rapidly after fertilization, hence telomere length will be reset in the offspring (83).

Telomere length needs to be maintained in stem cells

Given the above, the most simplistic model for the maintenance of lifespan through multiple generations, would be that telomere length is reset through the maintenance of long telomeres in germ cells and reset upon fertilization to then decrease continuously during the lifespan of an animal. However, this is an oversimplification, which became evident already when telomere length was first studied in mice. It was found that the decreasing telomere length differed greatly between different tissue types (88). A current hypothesis is that telomeres are maintained in adult stem cells, which resides in specialized compartments called stem cell niches. Stem cells are defined by the characteristic of not undergoing replicative senescence, hence being able to maintain telomere length.

Adult stem cells can divide asymmetrically to produce one daughter cell destined to differentiate into a tissue specific cell type that eventually loses expression, and can only divide a finite number of times before reaching senescence. The other daughter cell will remain a stem cell, capable of continuously replenishing the aging tissue with new cells. Imperfections in the ability to maintain these pools of stem cells could then explain organismal aging. Such stem cell niches have been characterized for a number of tissues. One of the best characterized is that of the hair follicle. Here, stem cells reside in a compartment called the bulge (95). They give rise to the rapidly dividing cells that make up hair and other follicular cells, but can also migrate to the basal layer of the epidermis and contribute to wound healing (96). Telomerase knock-out mice have fewer hair follicles with actively growing cells, resulting in baldness, and they show impaired wound healing (93). Bulge stem cells but not surrounding cells have telomerase activity, and as expected the stem cells have longer telomeres than the surrounding cells (97). Longer telomeres in adult stem cells compared to the more differentiated surrounding cells are a general feature of stem cell niches (97). Telomere length decreases with aging in the adult stem cells, as well as surrounding cells, suggesting that telomere lengthening in these compartments is limiting and contributes to aging (97).

ESCs have a rapid cell cycle (27), thus they are expected to require high telomerase activity. Indeed, telomerase knock-out in mESCs results in reduced telomere length and a corresponding defect in cell growth (98). In early development telomeres are rapidly lengthened from the zygote stage until the blastocyst stage, where ESCs are isolated from the ICM (83). mESCs may thus resemble a stage in development where long telomeres are maintained. Intriguingly, mESC telomeres have also been suggested to undergo occasional "super-lengthening" by recombination (99). Zscan4 was identified as a factor important for cell growth and telomere lengthening in mESCs (99). Normally it is only transiently expressed in a subset of mESCs in culture, but when overexpressed it co-localizes with telomeres, and a vast increase in Telomere Sister Chromatid Exchange (T-SCE) is observed (99). This occurs together with telomeric localization of proteins involved in recombination, but without other genomic instability (99). The existence of alternative telomere lengthening in mESCs is unlikely to only be an artifact of cell culture conditions, as recombination based telomere lengthening evident by Rad50 at telomeres has been observed in isolated embryos during the first divisions of the zygote, decreasing in the blastocyst (83). The importance of robust telomere lengthening and maintenance for pluripotent stem cells is highlighted by reduced efficiency when creating iPSCs from telomerase knock-out cells (100). Conversely, reprogramming efficiency increases when Zscan4 is added to the usual reprograming factors and telomere length is increased (101). When a cell is reprogrammed from a more differentiated state to an iPSC it requires extensive telomere lengthening (100). Thus, in terms of telomere length it is rejuvenated and its "molecular clock" is turned back.

Telomere lengthening in cancer

While stem cells have an important part in normal mammalian development, the ability of cells to divide a seemingly infinite number of times is also a hallmark of another cell type; cancer cells. As expected, most cancers over express telomerase (approximately 85% in total) (102). In most tumor types only a small fraction of tumors rely on ALT, however there are exceptions such as sarcomas and pancreatic neuroendocrine tumors that often rely on ALT (103,104). Inactivating mutations in the chromatin remodelers ATRX and DAXX are frequent in ALT tumors (103). Telomere length is often short compared to the tissue from which the cancer originate (89), however telomere length is maintained at this level. In tumors and cells lines relying on ALT telomere length is instead very heterogeneous (81). The lack of telomere length maintenance in most adult tissues can serve as an important tumor suppression mechanism.

Regeneration *versus* cancer

Laboratory mice may not be the best model organism for studying the significance of telomere lengthening in cancer. They have much longer telomeres than humans and cells from mice unlike humans can spontaneously immortalize in culture, even after telomerase knock-out (92). Still, telomerase knock-out mice show decreased incidence of skin cancer (105), while transgenic mice that over express telomerase in basal keratinocytes have greater incidence of skin cancer (106). Telomerase over expressing mice also show fewer signs of aging, increased wound healing, and increased maximum lifespan, i.e. the opposite phenotype of telomerase knock-out, suggesting that telomerase can be a target for regenerative therapies, however this comes at the expense of increased incidence of cancer (106,107). These mice only show a very small increase in average telomere length compared to wild type, indicating that telomerase could have a mitogenic effect other than simply preventing critically short telomeres (107). A mouse model was later developed that also over expressed telomerase, while having cancer resistance due to over expression of p53, p16 and p19 ARF (108). In this model the life span is increased to an even greater extent (108). The length of individual telomeres of the cells in the hair follicle was also studied, revealing that in this system telomerase over expression does increase the average telomere length and reduces the number of short telomeres (108).

Telomerase, its assembly and recruitment to telomeres

Telomerase is a reverse transcriptase that carries its own RNA template. The core enzyme is called telomerase, consisting of the protein telomerase reverse transcriptase (TERT) and the telomerase RNA component (TR or TERC) (109,110). Telomeres are lengthened by TERC hybridizing to the telomeric ssDNA and TERT adding repeats of 5’-GGTTAG-3’, followed by multiple rounds of translocation and synthesis (111).

TERT together with TERC are the essential components of the catalytically active enzyme. Active human telomerase is a dimer (112), and recently it was reported that two of the most commonly used human cell lines, HEK 293T and HeLa, contain active telomerase corresponding to about 240 telomerase monomers per cell (113). This suggests that in S-phase, when telomerase associates with and lengthens telomeres (114,115), there will be fewer telomerase dimers than there are telomeres. Hence, telomerase levels can be limiting to telomere lengthening. Overexpression of either one of the two essential components leads to more active telomerase being formed, but there also seems to be a pool of unassembled telomerase (113). Intriguingly, telomerase seems to be able to find the shortest telomeres, and thus selectively lengthens those telomeres that need it the most. This has been

most convincingly shown in yeast, where genetic techniques exist for shortening a single telomere selectively (116). Evidence for such selective lengthening also exists in mammals. In heterozygous TERT knock-out mESCs telomere length is initially reduced, but homeostasis is reached after a few population doublings and telomere length is maintained at this length without the cells losing their immortality or becoming genetically unstable (117). Hence, it seems that while the lower amounts of telomerase are insufficient for maintaining very long telomeres, the remaining functional telomerase prevents replicative senescence by lengthening the shortest telomeres. This suggests that the assembly of telomerase, and its ability to associate with telomeres are imperative for its ability to maintain telomere length homeostasis. The actual activity of the enzyme however, is well in excess of what is required (113).

Consistent with this notion, telomerase exists in cells as a large holoenzyme, encompassing several additional proteins which have roles either in the assembly, stability or transport of telomerase (118,119). Dyskerin (DKC1) is perhaps the most well studied, and together with its interaction partners pontin and reptin, is required for telomerase assembly and maintaining TERC stability (90,120). DKC1 interacts with the H/ACA motif of TERC, while pontin-reptin link DKC1 to TERT (90,120). In contrast to DKC1, immunodepletion of pontin-reptin does not abolish telomerase activity in extracts, suggesting that unlike DKC1 they are not an essential part of the active enzyme *in vivo* (120).

Similar to DKC1, another protein called telomerase Cajal body protein 1 (TCAB1), was found to be part of all active telomerase in cell extract (121). It interacts with TERT, TERC and DKC1, but rather than being required for telomerase assembly, it is required for telomerase localization to Cajal bodies and proper telomere lengthening in human cells (121). Later it was shown that the Cajal body as such is dispensable for the action of TCAB1 in facilitating the recruitment of telomerase to telomeres (122). The exact mechanism of TCAB1-mediated telomerase recruitment is unknown. One would expect a telomerase recruitment factor to be a protein that interacts both with telomerase and telomeric DNA (directly or via another protein) in order to bring the enzyme in direct contact with its substrate, but the latter activity has not been shown for TCAB1.

Other proteins are however known to do this. The most well studied telomerase recruitment factor is TPP1. It interacts directly with telomerase via a small patch of exposed amino acids, called the TEL patch (123), while its binding partner POT1 interacts directly with telomeric ssDNA via OB-

folds (52). In doing so it acts as a telomerase processivity factor, while also having a role in the recruitment of telomerase to telomeres (123,124). Surprisingly, POT1 is not required for telomerase recruitment, but the other TPP1 interaction partner TIN2 is (124). Hence, POT1 seems to mediate telomere lengthening when telomerase has been recruited to telomeres, while TPP1 and TIN2 are needed for telomerase to find the telomere.

Another well studied protein that interacts both with telomerase and telomeric ssDNA is hnRNP A1 and its proteolytic derivate UP1 (125). hnRNP A1 is a positive regulator of telomere length in cells (125), and increases the activity of telomerase (126). The UP1 portion of hnRNP A1 can pull-down telomerase from cell extracts, possibly due to its binding to TERC (127). However, while immunodepletion of hnRNP A1 from cell extracts decreases telomerase activity it does not deplete TERC levels and adding back hnRNP A1 alone restores activity (126). This suggests that the effect of hnRNP A1 on telomerase activity is separate from its interaction with telomerase. How hnRNP A1 affects telomerase activity is not fully understood. It was first reported that hnRNP A1 increases telomerase activity in a concentration dependent manner (126). A later study suggested that while hnRNP A1 increases telomerase activity it decreases processivity and above a threshold concentration, has an inhibitory effect on activity (127). A recent study encompassing more extensive biochemical work showed that hnRNPA1 does not increase telomerase activity *per se*, rather it can bind to and stop the inhibitory effect TERRA has on telomerase, while if present in excess can itself inhibit telomerase by blocking its substrate (128). The discrepancies between these observations may be due to the former two studies using the telomeric repeat amplification protocol (TRAP) assay with cell extracts (126,127), while the later used a direct telomerase assay and purified telomerase (128). It is possible that hnRNP A1 has different effects depending on holoenzyme composition, assembly and stability, which it may itself play a part in regulating, even though initial reports did not suggest this (125). Future studies addressing the effects of hnRNP A1 on telomerase biogenesis, its interaction with TERC, TERRA and telomeric ssDNA, as well as its effect on telomerase recruitment, may help shed light on its role in telomere length regulation.

Alternative DNA structures affect telomere accessibility

As mentioned above several factors are involved in aiding telomerase in finding its substrate, but there are also factors that promote the action of telomerase when at telomeres. As mentioned previously, telomerase activity *per se* is not expected to be limiting to telomere lengthening *in vivo*. This is evident from telomerase activity in biochemical assays being high compared

to the required lengthening needed to compensate for telomere loss during each cell division (113,129). However, a potential hindrance to telomere lengthening by telomerase *in vivo* is the formation of alternative DNA structures that obscure the 3' terminus, which must be available for TERC hybridization. Structures that may do this include the previously mentioned T-loops and G-quadruplexes.

The formation of a T-loop may result in the 3' terminus being hidden due to the invasion of the 3' overhang on the double-stranded part of the telomere. Little biochemical evidence exists for this, since the primer used in telomerase activity assays is usually a short oligonucleotide that is unable to form such a structure. Indirect evidence does however support such a model for telomerase inhibition. This stems mainly from the fact that TRF2 promotes the formation of T-loops (67,69), while being a negative regulator of telomere length (130).

The formation of a G-quadruplex at the 3' terminus may also affect telomerase binding and activity. Commonly, telomerase activity assays use primers that do not form G-quadruplexes, such as the short TS primer that does not contain the endogenous telomere sequence (131). Shortly after the discovery that Oxytricha telomeres forms G-quadruplexes, an assay comparing primers corresponding to different telomeric tract lengths concluded that telomerase activity is severely compromised when a four-repeat primer able to form an intramolecular G-quadruplex is used compared with shorter primers (132). When the shorter primers were lengthened enough to allow formation of an intramolecular G-quadruplex, it resulted in increased telomerase dissociation (132). Similar results were later obtained with human telomerase when formation of intramolecular G-quadruplexes were promoted by K^+ in reactions extending a three-repeat primer (133). Activity also decreases when human telomerase is unable to bind a G-quadruplex forming primer containing four repeats compared to a non-G-quadruplex forming primer, and the inhibitory effect of small molecule G-quadruplex ligands increases in this setting (134). Similarly, some (but not all) G-quadruplex ligands decrease telomerase processivity (134), as does molecular crowding agent PEG200 that promotes G-quadruplex formation (135). All G-quadruplexes are not equally detrimental to telomerase activity. Tetrahymena and Euplotes telomerase extend gel-purified intermolecular G-quadruplex primers relatively well compared to intramolecular G-quadruplexes, but not as well as the unfolded primer (136). Tetrahymena TERT binds directly to the intermolecular, but not the intramolecular, G-quadruplex, albeit not with as high affinity as for the unfolded conformation

(136). Similarly, the Est1p subunit of *S. cerevisiae* telomerase promotes formation of an intermolecular G-quadruplex (137).

Biochemical experiments have clearly shown that G-quadruplexes can affect telomere lengthening by telomerase, but is this also the case *in vivo*? The most comprehensive evidence for this comes from lower eukaryotes. The fact that both ciliate and yeast telomerase interacts with intermolecular G-quadruplexes strongly suggests they do exist during telomere lengthening *in vivo*. In support of this, the G-quadruplex promoting activity of Est1p is required for its role as a positive regulator of telomere length *in vivo*, possibly through aiding telomerase translocation (137). In another ciliate, *S. lemnae*, TEBPβ stabilizes a G-quadruplex and recruits telomerase to telomeres following phosphorylation in S-phase, which results in telomerase dependent displacement of TEBPβ and unfolding of the G-quadruplex (138,139). The observations made in *S. lemnae* are especially compelling considering that G-quadruplexes can be visualized *in situ* in macronuclei (138,139). There is indirect evidence for the importance of G-quadruplexes in affecting telomere lengthening also by human telomerase. The UP1 fragment of hnRNP A1, that increases telomerase activity, also promotes the resolution of a secondary structure presumed to be a G-quadruplex (126). hnRNP A2*, a splice variant of hnRNP A2 with a deletion in the glycine-rich domain, resolves telomeric G-quadruplexes and promotes telomerase activity and processivity, interacts with TERC and colocalizes with telomerase in Cajal bodies (140). As previously mentioned, POT1-TPP1 both resolves G-quadruplexes and increases telomerase processivity (74,141). Given the impact of G-quadruplex formation on telomerase processivity (133), it seems likely these functions are connected. Surprisingly, when a mutated telomerase incorporating repeats of TTAGGC instead of TTAGGG was used (thus being unable to form a G-quadruplex), POT1-TPP1 still had an effect on processivity as long as the primer contained an upstream POT1-TPP1 binding site (142). Thus the initial binding of POT1-TPP1 at a site 5'-distal to the telomerase binding site is enough to promote processivity. However, since this system also abolishes the need for G-quadruplex resolution it cannot be excluded that POT1-TPP1 also promotes wild type telomerase processivity by resolving G-quadruplexes. In addition to proteins that both have roles in G-quadruplex metabolism and telomere lengthening by telomerase, further proof that G-quadruplexes have a role in regulating telomerase comes from small molecule G-quadruplex ligands. Some of the ligands that affect telomerase activity in biochemical experiments have also been shown to induce telomere shortening in human cancer cell lines (143-145). However, ligands such as telomestatin also prevent proteins such as POT1 and TRF2 from binding telomeric DNA (76,146). This makes it difficult to discriminate

between direct effects on telomerase and indirect effects caused by interference with other telomere associated proteins.

In summary, it is clear that G-quadruplexes play a role in regulating telomere lengthening by telomerase, but their exact role in regulating telomere length homeostasis *in vivo* is not as well established and warrants further investigation. For instance, in *S. lemnae* and *S. cerevisiae* telomeric G-quadruplex formation and lengthening by telomerase are cell cycle dependent (137,139). In human cells, G-quadruplexes form more readily during S-phase when telomere lengthening is known to take place (16), but it is unknown whether G-quadruplex formation increases also at telomeres specifically. If telomerase does indeed lengthen short telomeres selectively, can G-quadruplex formation be part of the mechanism regulating this? Less biochemical evidence exists for T-loops affecting telomerase activity, but it stands to reason that telomerase would be unable to extend the hidden 3’ end. As discussed previously the resolution of T-loops is imperative for maintaining telomere length, but seems to be related to deletion of telomeric circles rather than preventing telomere lengthening by telomerase (62). The T-loop might act more as an ON/OFF switch, while G-quadruplexes could have more of a fine tuning effect.

Telomeric heterochromatin regulates telomere length, and is affected by G-quadruplex binding proteins and telomeric transcripts

As previously mentioned, telomeres are heterochromatic. They are rich in di- and trimethylated histones (H3K9 and H4K20), leading to the recruitment of heterochromatin protein 1 homologs Cbx1, Cbx3 and Cbx5 (also known as HP1β, γ and α respectively) (147,148). The epigenetic status of telomeres are dependent on their length, as telomeres in MEFs from Terc knock-out mice lose heterochromatic marks as telomere length decreases (149). Conversely, loss of heterochromatic marks leads to telomere elongation (147,150).

In fruit flies and yeast it is well established that heterochromatin formation at telomeres leads to the silencing of nearby genes, termed the telomere positioning effect (TPE) (151,152). In human there are few genes close enough to the telomere to be affected by this, but it does affect expression of the *DUX4* gene that is involved in Facioscapulohumeral muscular dystrophy (FSHD) (153). However, the telomere is itself transcribed and telomere length has an effect on TERRA transcription (154,155). Expression levels of TERRA decrease with telomere length, but the length of each transcript increases, and the total amount of TERRA is higher at longer telomeres (155). TERRA localization to telomeres causes heterochromatin formation.

This is evident from them showing the same correlation with telomere length, by compromised heterochromatin formation upon TERRA depletion, and by having the same cell cycle dependency with successively decreasing levels through S-phase and G2/M (155,156). This may facilitate a negative feedback loop for TERRA expression, creating a self-regulating mechanism for maintaining the right chromatin compaction (155). Since heterochromatin also affects telomere length, it may also provide a mechanism for maintaining telomere length homeostasis. TERRA may also do this independently of heterochromatin by directly inhibiting telomerase (157).

TERRA associates with Cbx5/HP1α, ORC, and H3K9Me3 and may thus induce heterochromatinization by aiding their recruitment (156). TERRA contains UUAGGG repeats (154). When forming a G-quadruplex, TERRA binds the N-terminal GAR domain of TRF2 (158), and this domain is required for both TERRA and ORC localization to telomeres in cells (156,159). TRF2 can simultaneously bind double-stranded DNA through its Myb domain, and thus tether TERRA to the telomere (158). The protein FUS/TLS was also recently shown to simultaneously bind both TERRA and telomeric DNA G-quadruplexes through its RGG3 domain (160). FUS acts as a negative regulator of telomere length (160). FUS recruits the methyltransferase Suv4-20h to telomeres, leading to increased H3K20 tri-methylation (160) (figure 5). TERRA may also find the telomere by invading the dsDNA by hybridizing with the C-rich strand and displacing the G-rich strand, forming an R-loop. To what extent such structures exist at

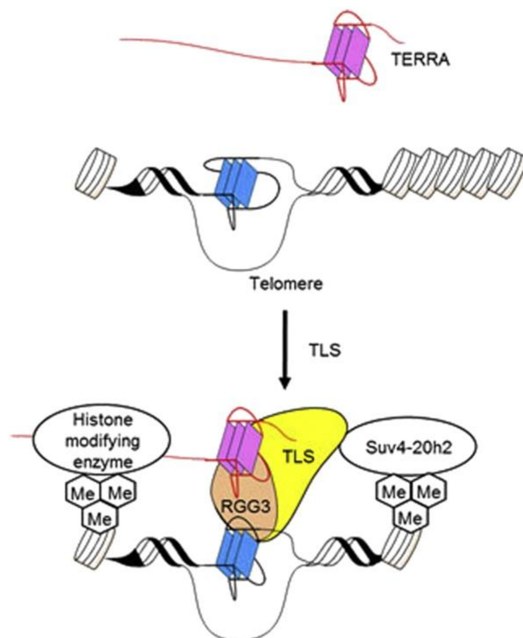

Figure 5. Proposed model for how FUS/TLS may bind a TERRA G-quadruplex (red and pink) and telomeric DNA G-quadruplex (black and blue) simultanously to recruit histone methyltransferase Suv4-20h2 and possibly other histone modifying enzymes. As suggested in the text, TERRA might also hybridize to the displaced C-strand. Reprinted from Chemistry & Biology, Vol. 20, Takahama, K. *et al.*, "Regulation of Telomere Length by G-Quadruplex Telomere DNA- and TERRA-Binding Protein TLS/FUS", Pages 341-350, copyright 2013, with permission from Elsevier.

telomeres and their possible role in telomere maintenance is poorly understood. Telomeric R-loops were first found to exist at yeast telomeres (161), and recently were shown to also exist in a number of human cancer cell lines (162). R-loops were found to affect telomere recombination in ALT cell lines (162). Telomeric R-loops may very well incorporate and be stabilized by G-quadruplexes. Hypothetically, these could be formed by the displaced G-strand, by protruding RNA, or as RNA:DNA hybrids. Indeed, in addition to binding both telomeric DNA and RNA G-quadruplexes, TRF2 also binds a hybrid RNA:DNA G-quadruplex *in vitro* (158). FUS was recently suggested to be involved in an R-loop mediated DNA damage response together with hnRNPU/SAF-A, suggesting that its function in telomere length regulation may also involve R-loops (163). ATRX, a chromatin remodeler in the SWI/SNF family, binds to telomeres via an interaction with histone variant H3.3 in mESCs, but has also been shown to interact directly with G-quadruplexes (164,165). ATRX protects telomeres from DDR, facilitates H3.3 and HP1α localization to telomeres, and represses TERRA levels in mESCs (164,166).

Effects of telomere subnuclear localization

Different properties of a cell are regulated by spatial organization into areas dedicated to certain processes. Spatio-temporal organization of chromatin may have functions in regulating nuclear processes, and thus also telomere maintenance. The nuclear matrix or scaffold is a structure on which the nuclear content is organized. The inner nuclear matrix connects to the nuclear lamina which lines the nuclear envelope. Some genomic DNA adheres tightly to the nuclear matrix. These domains are called Scaffold/Matrix Attachment Regions (S/MARs).

Yeast telomeres are localized to the nuclear envelope, which is linked to telomere silencing through heterochromatinization (167). In contrast, mammalian telomeres are found throughout the nucleus during interphase. Their localization does however vary with the cell cycle, with more telomeres localizing to the nuclear envelope in early G1 compared to the rest of interphase (168). Localization to the nuclear lamina is not a prerequisite for heterochromatin formation in human cells, as centromeres (which are heterochromatic) do not show this localization (168). Mammalian telomeres are known S/MARs, attaching to the matrix throughout the cell cycle (169,170). When introducing the telomeric sequence internally rather than at the end of a chromosome, it does not function as an S/MAR, suggesting the 3’ terminus is essential for this function (169). All telomeric proteins do not associate with the nuclear matrix to an equal extent, suggesting that the association of the telomere to the nuclear matrix and the composition of the

matrix are important for regulating telomere function. Shelterin protein TIN2 exists in a longer and shorter splice variant called TIN2L and TIN2S respectively (171). TIN2L is found to a greater extent in the nuclear matrix than other Shelterin proteins and its shorter variant (171). The hnRNP A2 splice variant hnRNP A2* also localizes to the nuclear matrix (140). As previously mentioned it is a positive regulator of telomere lengthening, by recruiting telomerease, and can unfold telomeric G-quadruplexes to promote telomerase processivity (140). Besides telomerase recruitment, spatial organization of telomeres may be important for heterochromatin formation and transcriptional regulation of TERRA expression, which as mentioned in the previous section also affects telomere length. This is evident from studies on telomeres after knock out of the *Lmna* gene that encodes A-type nuclear lamins (172). A-type lamins are a major constituent of the nuclear lamina that lines the nuclear envelope, but may also be part of an internal nucleoskeleton (173). When *Lmna* is knocked out in MEFs it leads to an altered distribution of telomeres and telomere shortening (172). These cells also have less H3K20Me3 and lower levels of TERRA, suggesting that the decreased telomere length is linked to transcriptional regulation (172). *Lmna* knock-out cells also have defects in processing of dysfunctional telomeres through NHEJ (172).

3’ overhang length regulation and the role of G-quadruplexes

Several telomere binding proteins essential for telomere maintenance bind single-stranded telomeric DNA, and the 3’ overhang acts as the DNA primer for telomerase. Hence, it is not surprising that maintenance of the 3’ overhang is important for telomere function. The lagging strand telomere will carry a 3’ overhang by default after replication, since the DNA polymerase α-primase complex cannot ensure priming to the most distal end of the telomere and because there is no mechanism for replacing the distal RNA primer with DNA (174). The leading strand on the other hand will be blunt ended (174). In cultured cells lacking a telomere lengthening mechanism (i.e. telomerase or ALT) overhang length is proportional to the rate of telomere loss with each population doubling, and hence related to the rate of cellular senescence (175).

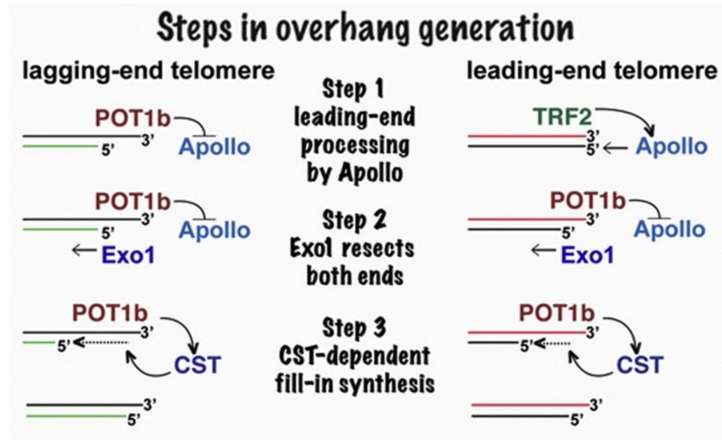

Figure 6. Summary of 5' end resection and overhang length regulation at mammalian telomeres. Reprinted from Cell, Vol 150, Wu, P. *et al*., "Telomeric 3' Overhangs Derive from Resection by Exo1 and Apollo and Fill-In by POT1b-Associated CST", Pages 39-52, Copyright 2012, with permission from Elsevier.

There needs to be a mechanism in place for resection of the 5' strand. The 5'→3' exonuclease Apollo was identified as controlling overhang length and acting specifically at leading strand telomeres (176,177). TRF2 interacts directly with and recruits Apollo (178,179), while POT1 (specifically POT1b in mouse) instead inhibits Apollo (180). Both leading and lagging strand telomeres have the same 5' terminal sequence, suggesting that there is controlled resection also at the lagging strand (181). Another 5'→3' exonuclease, Exo1, acts on both leading and lagging strand telomeres (180). The human homolog of the *S. cerevisiae* CST (Ctc1-Stn1-Ten1) complex protects telomeres from DDR by regulating overhang length by promoting C-strand fill-in (180,182-184). Figure 6 summarizes the current understanding of telomere overhang regulation.

In addition to the above mentioned proteins, G-quadruplexes may also have a role in regulating 5' end resection. In yeast a G-quadruplex stabilizing and a G-quadruplex resolving helicase, Stm1 and Sgs1 respectively, have been implicated in the regulation of overhang length (185). Stm1 rescues while Sgs1 exacerbates the phenotype of a temperature sensitive Cdc13 mutant (the yeast homolog of Ctc1 in the CST complex) (185). In an effort to elucidate if G-quadruplexes contribute to telomere maintenance in yeast, the same temperature sensitive Cdc13 mutant was used together with various strategies for G-quadruplex stabilization (by the above mentioned Stm1 and other G-quadruplex binding proteins, Sgs1 deletion, and by G-quadruplex stabilizing ligands), showing that they do indeed rescue telomere maintenance by

inhibiting hyper-resection of the C-strand (186). This effect was not seen in mutants carrying a telomerase that adds a mutant repeat unable to form G-quadruplexes (186). It was demonstrated through biochemical experiments that G-quadruplex stabilization inhibits 5' end resection by yeast Exo1 (186). A model for how G-quadruplex stabilization contributes to overhang length regulation is shown in figure 7.

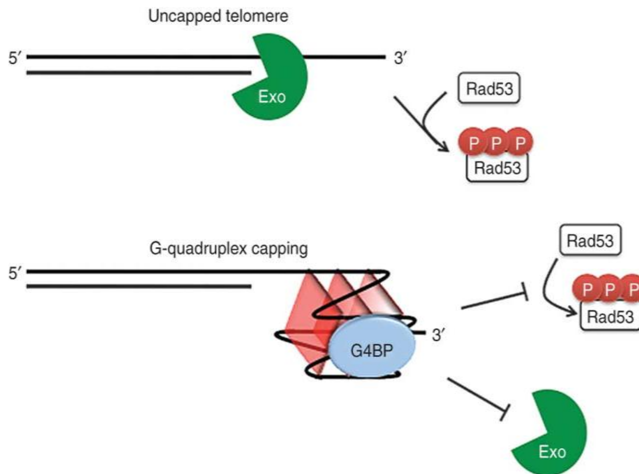

Figure 7. Mechanism for G-quadruplex mediated telomere protection in yeast by a G-quadruplex binding protein (G4BP). Reprinted by permission from Macmillan Publishers Ltd: Nature Structural and Molecular Biology (Smith, J.S. *et al.*, "Rudimentary G-quadruplex-based telomere capping in Saccharomyces cerevisiae", vol. 18, p. 478-485), copyright (2011).

1.3.4 Telomere maintenance as a whole

While this text has focused on end-protection and end-replication individually, it is important to note that these are by no means separate mechanisms. It is only logical that for any telomere binding protein to function, there must be telomeric DNA for it to bind to (and in some cases a sufficiently long 3' overhang), hence telomere length is obviously related to end-protection. The consequence of faltering end-protection and end-replication are similar, leading to activation of the DDR and subsequent cell death or senescence (39). Hence it is not surprising that many of the same proteins have essential roles in mediating end-protection and in regulating telomere lengthening. This is the case for the Shelterin complex, which serving as the core telomere-associated protein complex plays a role in basically all known telomeric processes. Individual complex components and protein domains within components can have separate functions that may be related to one process but not another. The latter is the case for TPP1, where

enough is known about its structural-functional properties to create mutants defective in telomerase recruitment without affecting its direct role in inhibiting the DDR (123). Such detailed knowledge is important, as it could aid in the development of cancer therapies targeting telomerase function. Experiments in mice suggest telomere maintenance as a whole is the important aspect of telomeres in aging. TRF1 knock-out mice have telomeres of normal length, but defective end-protection leads to telomere fragility and fusions in MEFs (187). Mice with conditional TRF1 knock-out in stratified epithelia die perinatally, lack hair follicles, and show skin hyperpigmentation (187). When the perinatal lethality is rescued by p53 deletion it leads to increased incidence of cancer, probably owing to the observed genetic instability, but also to epithelial abnormalities similar to dyskeratosis congenita (187). This shows that aging phenotypes can be the result not only of telomere shortening, but rather that telomere dysfunction in general might be an important factor in aging and cancer.

1.4 Transcriptional regulation

All cells within a species carry the same genetic information encoded in their DNA (with the exception of some cells of the immune system). How can cells with the same genome carry out vastly different functions in the adult body and during different stages of development? This is made possible by differential gene expression, meaning that genes important for certain functions are expressed at different levels in different cell types. Gene expression starts with the DNA of genes being transcribed into RNA through the process of transcription. The RNA transcript is then modified and exported out of the nucleus to be translated into a protein which carries out a biological function (e.g. a metabolic enzyme or structural component of the cell), or in some cases the RNA is non-coding and instead carries out its biological function without being translated (e.g. the previously mentioned TERC and TERRA). Gene expression can be regulated at all levels between the start of transcription until the breakdown of the encoded protein. Here I will focus on the process of transcription.

Transcription of genes encoding proteins is exclusively carried out by RNA polymerase (Pol, or RNAP) II, while Pol I and III carry out transcription of ribosomal RNA, transfer RNA and some other non-coding RNA. This text will focus on Pol II mediated transcription, but the principles are the same for all three. Transcription is divided into three steps:

During *initiation* the transcription machinery is set up on a gene promoter that lies upstream of a transcription start site (TSS). Pol II is a large complex

consisting of 12 subunits (Rbp1 to 12). Pol II forms a pre-initiation complex with a set of general transcription factors (TFIIB, D, E, F and H) that recognize sequences in the promoter and help form a transcription bubble by ATP dependent unwinding of the dsDNA making the ssDNA template available (188).

Elongation starts with the Pol II being cleared from the promoter. It forms a locked clamp around the DNA (189), a more stable transcription bubble is formed, allowing the polymerase to slide along the gene synthesizing RNA in a processive manner, meaning one Pol II binds and synthesizes the entire length of the transcript.

Termination of transcription occurs when Pol II reaches a termination sequence. Some of the factors that control this are the same as those controlling the mRNA 3' end-processing, including cleavage and polyadenylation specificity factor (CPSF) and cleavage and stimulation factor (CstF) (190,191).

The transition from the different steps of transcription is mediated by phosphorylation of the C-terminal domain (CTD) of the large Pol II subunit Rpb1, which results in a series of protein interactions controlling the action of the polymerase (192). It is worth noting that while transcription is usually explained in this manner of discrete steps, there is actually considerable overlap, as evident by factors involved in splicing, capping, and termination associating with the Pol II already during initiation (192).

The rate of transcription is an important regulator for gene expression. Transcription factors bind sequences specific to certain genes and can either activate or repress transcriptional initiation. hnRNP U/SAF-A that is the topic of paper I and II of this thesis acts as a transcription factor by binding sequences in the promoter of certain genes, such as the Oct4 proximal promoter (193), and also has a role in transcriptional elongation through inhibition of Pol II CTD phosphorylation by TFIIH (194).

1.4.1 Chromatin and its remodeling

If all DNA in a human cell was stretched out it would measure around 2 meters long in total. To make it fit into the cell nucleus eukaryotic DNA is packed into chromatin, making it much more compact. The core units of chromatin are nucleosomes. They consist of 146 bp of DNA wrapped as a two-turn superhelix around eight histones (two copies each of H2A, H2B, H3 and H4) (195). Between nucleosomes lie short segments of linker DNA, which are bound by linker histone H1. Nucleosomes are further arranged into

10 nm "beads on a string" fibers, which can be further packed into a 30 nm solenoid or zigzag structure (196,197). These secondary structures can be further arranged into less well defined higher order (tertiary) structures. The structure of chromatin on all levels, starting with alternative nucleosome compositions, is variable and highly dynamic (198). The arrangement of DNA into chromatin serves not only to accommodate DNA into the limited space of the nucleus, but also to regulate essentially all processes involving DNA, i.e. replication, transcription, repair and recombination. Chromatin brings order to the large eukaryotic genomes, bringing genomic loci that are separated by a long distance in terms of sequence into close proximity. The tight packing of the DNA around histones will also leave it inaccessible to other proteins, including the transcription machinery.

As previously mentioned, the compaction of chromatin varies. Very compacted regions are referred to as heterochromatic, while more loosely packed chromatin is called euchromatic. The former is transcriptionally silent (although this is not strictly true, as evident e.g. from telomeres being transcribed) and the latter is transcriptionally more active. Chromatin compaction is regulated by enzymes that posttranslationally modifies histone tails (199). The most well studied modifications are methylation and acetylation, governed by three classes of proteins; histone methyltransferases (HMTs), histone acetyltransferases (HATs) and histone deacetylases (HDACs). Histone tails point out from the nucleosome and may regulate both DNA-histone interactions and histone-histone interactions that mediate chromatin compaction (195). As already mentioned, trimethylation of Lys20 on H4 (H4K20me3) is a signature of heterochromatin. This modification promotes nucleosome interactions and formation of compact fibers (200). Acetylation of Lys16 on the same histone tail conversely destabilizes nucleosome-nucleosome interactions, leading to a less compact structure (201). Post-translational modifications also lead to recruitment of other proteins that affect chromatin compaction, such as HP1 proteins that interacts directly with methylated Lys9 of H3 (202).

ATP-driven chromatin remodeling complexes can alter the packaging of chromatin by changing the position or structure of nucleosomes. This will make the DNA accessible to other proteins such as transcriptional activators and repressors, but chromatin remodeling is important for all process involving chromatin (203). The ATPase domain that consumes ATP to break DNA contacts with histones is conserved between remodelers, while the flanking domains that mediate their interactions with other proteins vary and divide remodelers in different subfamilies (204). The different subfamilies have different functions and act in different processes. They are large

complexes with many components, which may also change between cell types to mediate specific functions (203). There are two human SWI/SNF complexes; PBAF and BAF. BRG1 (Brahma related gene 1) is the ATPase in PBAF, while both BRG1 and BRM (Brahma) can function in BAF (205). BRG1 binds nucleosomes via its bromodomain that specifically interacts with acetylated histone H3 and H4 tails (206,207). Histone acetylation aids in SWI/SNFs ability to induce sliding of nucleosomes (208). Human SWI/SNF plays important roles in regulating transcriptional activation. Disruption of nucleosomes by human SWI/SNF aids in transcription factor binding (209). It also associates with gene promoters and interacts with transcription factors (210), and BRG1 associates with a Pol II holoenzyme (211). BRG1 can also aid transcriptional elongation by helping Pol II overcome nucleosomal barriers (212). An mESC specific BAF complex exists (esBAF), with a specific subunit composition that is built on BRG1 rather than BRM (213). esBAF function is essential for maintaining mESC pluripotency and self-renewal (213), having a role in both transcriptional regulation and replication (214-216). BRG1 is the topic of paper II of this thesis.

1.4.2 The role of G-quadruplexes in transcriptional regulation

G-quadruplexes may have widespread and important roles in regulating gene expression. Bioinformatic approaches have shown that putative quadruplex forming sequences (PQSs) are commonly found upstream of the TSS at promoters (14,15). The exact number varies with the search algorithm used and how far from the TSS it searches, but around >40% of the total TSSs in human have at least one PQS in the 1 kb upstream of the TSS (15). Genes containing PQSs in promoters are conserved in orthologs of human, mouse and rat, and genes with PQSs in their promoters are enriched for tissue specific expression, further supporting that they are important gene regulatory elements (217). Recently it was shown *in vitro* that the negative supercoiling caused at the back of a transcribing polymerase promotes G-quadruplex formation (218). Thus, transcription might cause a signal that is relayed over long genomic distances, and then sensed by factors recognizing G-quadruplexes formed in gene regulatory elements. This adds further support to the PQS of gene promoters being relevant in transcriptional regulation.

Several promoter sequences have been shown to form G-quadruplexes in biochemical and biophysical experiments, but less is known about their function *in vivo*. The most insight of how G-quadruplexes may regulate transcriptional activity probably comes from the c-Myc promoter and the

protein nucleolin (Ncl). The c-Myc promoter contains a nuclease hypersensitive element (NHE III_1) where a G-quadruplex can form (13). This G-quadruplex is bound and stabilized by Ncl, and when doing so it represses transcription (219). The same area of the c-Myc promoter is bound by NM23H2, which acts as transcriptional activator (220). Unlike Ncl it resolves the G-quadruplex and binds the ssDNA in an extended conformation (221). Thus these two proteins with opposing activities on G-quadruplex stability have opposing roles in c-Myc transcriptional regulation. Inserting the NHE III_1 sequence in a negative supercoiled plasmid promotes G-quadruplex formation, lending further support to the hypothesis that negative supercoiling may be a common mechanism for G-quadruplexes in regulating transcription (222). Transcription factor SP1 was also recently shown to bind G-quadruplexes (223). The SP1 minimal consensus motif (5'-GGGCGG-3') is prevalent in gene promoters, but while it is G-rich it is not predicted to form a G-quadruplex on its own, and SP1 binds to dsDNA *in vitro* (224). An SP1 consensus motif was identified in a sequence of the c-KIT promoter that also forms a G-quadruplex, and SP1 was found to bind the G-quadruplex with higher affinity than the dsDNA of the same promoter sequence (223). When analyzing SP1 ChIP on Chip data, sequences overlapped better with predicted quadruplex forming sequences than they did with the classical SP1 consensus motif, indicating that G-quadruplexes may be important for transcriptional regulation by SP1 (223).

Studies of Ncl and nucleophosmin (Npm1) also show another possible role for G-quadruplexes in transcriptional regulation; regulating the subnuclear localization of transcription factors (225). As the names suggests, Ncl and Npm1 normally reside in nucleoli, the specialized nuclear compartments where ribosomal RNA is transcribed and modified. However, as evident by Ncl's role at the Pol II transcribed c-Myc gene, and as shown in paper III of this thesis, Ncl has functions outside of nucleoli. The same is true for Npm1, and both proteins have been shown to shuttle in and out of nucleoli and between nucleus and cytoplasm following phosphorylation (226-229). Both Ncl and Npm1 bind G-quadruplexes in rDNA (225,230). Intriguingly, mutations in the basic C-terminal domain of Npm1 that binds G-quadruplexes are found in acute myeloid leukemia, where Npm1 is miss-localized from nucleoli (225). G-quadruplex ligands that disrupt Ncl and Npm1 rDNA G-quadruplex binding were shown to exclude them from nucleoli (225,230). Whether rDNA G-quadruplex binding also regulates e.g. the cell cycle dependent localization and functions of these proteins remains unanswered, but it is an intriguing possibility.

G-quadruplexes may also be involved in transcriptional regulation at the level of chromatin remodeling. ATRX is an ATP driven chromatin remodeler in the SWI/SNF family. It is mutated in alpha-thalassemia X-linked mental retardation syndrome (231). ATRX regulates gene expression by targeting tandem repeat sequences that are G-rich (165). DNA binding assays showed that ATRX interacts with a G-rich sequence when it forms a G-quadruplex, suggesting that G-quadruplex formation helps guide chromatin remodeling by ATRX (165).

2 AIMS

The aim of the first paper of this thesis was to elucidate the role of hnRNP U/SAF-A in telomere biology, and whether it involves G-quadruplexes.

The aim of the other two papers was to identify and characterize important protein-protein interactions in ES cells. In paper II the role of hnRNP U/SAF-A in global transcription together with the ATP-driven chromatin remodeler BRG1 was investigated, and in paper III interaction partners of Ncl were studied.

Since the work revealed hnRNP U/SAF-A as a G-quadruplex binding protein, and since Ncl has a known role in transcription by interacting with G-quadruplexes, this thesis discusses the possible role of G-quadruplexes in the processes of transcriptional regulation and telomere maintenance.

3 ASPECTS OF METHODOLOGY

In the work presented here many different techniques have been used, each being detailed in the methods section of the respective papers. Rather than reiterating details this chapter will discuss the methods used in a more general fashion categorized by what the assays are used for.

3.1 Studying protein-DNA interactions

The methods for studying protein-DNA interactions in this thesis can be categorized as either biochemical experiments where purified protein or extracts and their binding to DNA are being studied in the test tube, or methods detecting the association of proteins with DNA in the cell.

3.1.1 DNA pull-down

For studying protein-DNA interactions in the test tube, the assay used in paper I was DNA pull-down. In this assay biotinylated oligonucleotides are bound to streptavidin coated magnetic beads. The biotin-streptavidin interaction is one of the strongest interactions known. Magnetic beads make for convenient washing, as the beads can be held in place by a magnet as the buffer is changed. Following immobilization of oligonucleotides an extract or purified protein is added and incubated with the beads to allow binding. This is followed by washes and elution, e.g. by denaturation of bound proteins by boiling in SDS-sample buffer. The eluate is then run on a denaturing polyacrylamide gel, followed by Western blotting to detect proteins, i.e. they are transferred from the gel to a membrane and then detected using antibodies specifically recognizing the protein of interest. The main advantages of this method are its ease of use and the ability to study the behavior of proteins in complex mixtures (e.g. a cell extract). Immobilizing the DNA also allows for incubating it with different proteins or extracts sequentially, to gain insight of how one protein may interact with a preformed protein-DNA complex. Other varieties of this method exist where the oligonucleotide is immobilized on different surfaces, such as the wells of a micro-titer plate, which allows for other means of detection.

Another popular method for studying protein-DNA interaction, that was not used here, is electrophoretic mobility shift assay (EMSA). In this method proteins are mixed with labeled oligonucleotides, allowed to bind, and are then electrophoresed on a native gel. Protein bound DNA is then recognized as a shifted band, due to slower migration. Advantages of EMSA compared

to DNA pull-down is that you can get additional information about complex stoichiometry and conformation, e.g. you may have two shifts, one being a 1:1 DNA:protein complex and one being 1:2, or you may have two bands representing free probe in two different conformations and only see one being shifted. A disadvantage compared to DNA pull-down is that it is more difficult to obtain information about to what extent different proteins in a complex mixture binds to a DNA substrate. You can use an extract and then test if one protein is found in a shifted band by using an antibody to super shift it, but this can be more challenging than using DNA pull-down, where only one binding reaction per substrate is needed for studying multiple proteins at a time.

3.1.2 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) is a method used for detecting the association of a protein with DNA in cells. An extract is made that contains sheared chromatin (usually by sonication) to a desired size (which dictates the resolution of the assay in terms of "genetic distance"). Usually proteins are cross-linked by one or several agents before cell lysis to stabilize protein-chromatin binding. An antibody directed towards the protein of interest is then used for "precipitation", meaning that the protein is bound to the antibody and the antibody-protein-DNA complex is then immobilized on beads. The beads are then washed and the bound DNA is eluted. Which DNA sequences and to what extent they are bound is then analyzed. This can be done in multiple ways. For abundant repetitive sequences, such as telomeres, a convenient and robust detection method is to do a Southern dot-blot. Eluted DNA is blotted on a nylon membrane using a vacuum scaffold. It is then detected using Southern hybridization of a labeled DNA (or RNA, or modified oligonucleotide) probe with the antisense sequence to the sequence of interest. This is the method used in paper I of this thesis. A more common method for detecting specific sequences is PCR, preferably quantitative (as described later). The recent revolution in microarray and sequencing techniques has allowed for the detection of virtually all DNA sequences bound by a protein.

3.1.3 Immunofluorescence and fluorescence *in situ* hybridization

To simultaneously detect where in a cell a protein and a DNA sequence is, fluorophore labeled antibodies and nucleic acids can be used, that recognize a specific protein and DNA sequence respectively. This is called immunofluorescence (IF) and fluorescence *in situ* hybridization (FISH) respectively. IF-FISH was used in paper I of this thesis. A fluorophore is a

molecule that emits light of a certain wavelength after being illuminated by light of another wavelength. By choosing one fluorophore that is excited and emits light of a lower wavelength, and another fluorophore with a higher wavelength, the two can be distinguished (e.g. by green and red light). Additionally, it is common to fluorescently label one of the defining structures of the cell. Commonly a fluorescent DNA label such as DAPI (4,6-diamino-2-phenylindole) that binds dsDNA is used. Cells are grown and fixated on a coverslip or microscope slide. They are then incubated with antibody and nucleic acid probe in separate steps due to their different binding dynamics, meaning different buffers need to be used to block unspecific interactions, allow binding of the probe, and to wash away unbound probe. Usually IF is carried out first, due to the harsher conditions needed for FISH. The fluorescent signal is observed using a fluorescence microscope. In the work presented in this thesis confocal microscopy was used. With this system each fluorophore is visualized individually by excitation using separate lasers, and filters are chosen to only detect the light emitted from the particular fluorophore that is used. A confocal microscope is able to capture light from narrow optical planes by using a pinhole to eliminate out of focus signal. By taking images sequentially in a stack in the Z-direction, the assembled images can be used to generate a three-dimensional image. This method can thus be used to gain information of where a protein and a DNA sequence are located in the cell. If the two signals overlap it means the protein is in close proximity to the DNA sequence of interest. The resolution of a fluorescence microscope is however limited (to about 200 nm in the xy-direction and 500 nm in the z-direction), meaning one cannot be certain the protein is directly associated to the DNA of interest, or if it is associated with a very nearby sequence and whether or not it is a direct interaction.

3.2 Studying protein-protein interactions

The methods used for studying protein-protein interactions are similar in principle to the methods used for protein-DNA interactions. The only difference is that both of the interacting molecules are proteins. Below follows a brief description of the methods used in the work presented in this thesis. Many other methods do however exist, including biochemical experiments that address how proteins interact with each other and the affinity of the interaction, which has not been thoroughly addressed here.

3.2.1 Protein affinity purification

Protein affinity purification can be used to identify which proteins in an extract are interacting with a protein of interest. The protein of interest is conjugated to a solid support, which is usually made up of porous particles with a big surface area. These are then incubated with a cell extract and packed in a column, washed, and proteins are then eluted. Elution is usually carried out under increasingly stringent conditions, often by increasing the salt concentration of the elution buffer. High salt disrupts protein interactions. The eluate can then be run on a denaturing polyacrylamide gel to separate the proteins by size. Bands can be cut out from the gel, and the contained proteins identified by tandem mass spectrometry. This method was used in paper III of this thesis.

Another common strategy is to overexpress the protein of interest coupled to an affinity tag. The tag is often a peptide sequence which binds with high affinity to an antibody, but tags that bind other types of molecules can also be used. A lysate is then made that contains the protein in complex with its interaction partners, and these are then bound to a solid support that is conjugated to the antibody/molecule that binds the tag. The subsequent steps are similar to what is outlined above. This method has the advantage of allowing the capture of complexes formed in the living cell.

3.2.2 Co-immunoprecipitation

Immunopreciptation was briefly outlined in the description of ChIP above. This step is essentially carried out in the same manner for co-immunoprecipitation (co-IP), except that the extract can be made differently to fractionate proteins e.g. being cytoplasmic or nuclear. The main difference is that instead of analyzing the DNA that is bound by a protein of interest, the bound proteins are being analyzed. This is usually done by eluting proteins in high salt, acidic, or denaturing SDS loading buffer, followed by detection of the proteins by Western blotting. Co-IP was used in papers II and III of this thesis.

3.2.3 Immunofluorescence

This method is the same as outlined for IF coupled to FISH, only the FISH step is omitted. Instead two different antibodies are being used directed to the two proteins of interest. These antibodies are usually raised in two different species. Two secondary antibodies carrying different fluorophores and directed towards different species antibodies are then used to detect the primaries. The information gained is the same as in IF-FISH, except two

different proteins are detected, and the same limitations apply. IF is used in all of the papers presented in this thesis.

3.2.4 *In situ* proximity ligation assay

In situ proximity ligation assay (PLA) is similar to IF, except that only proteins in a proximity closer than 40 nm are detected (232). The initial steps of this method are identical to IF, except fluorophore labeled secondary antibodies are not used. Instead the secondary antibodies are conjugated to two separate oligonucleotides. When the secondary antibodies are close enough to each other the conjugated oligonucleotides can hybridize with added connector oligonucleotides and after ligation form a substrate for rolling circle amplification. A polymerase is added which amplifies the DNA. The amplified DNA is then hybridized to a fluorophore labeled oligonucleotide. The signal will be detected as discrete fluorescent dots where proteins are in close proximity. The advantages of this method over ordinary IF are that it enforces a more stringent requirement on proximity, and that it is easier to visualize and quantify. On the other hand you cannot detect information about the protein fractions that are not located in close proximity. PLA was used in all three papers presented in this thesis.

A popular technique for studying protein-protein interaction in cells, that was not used in the work presented in this thesis, is Förster resonance energy transfer (FRET), where proximity is instead detected by the ability of one fluorophore to transfer energy to a nearby fluorophore (233). By using fluorescently tagged proteins instead of antibodies, it is possible to visualize proteins and their interaction (or close proximity) in real time in living cells. An advantage with using PLA rather than FRET is that the endogenous rather than over expressed proteins are being studied. Several techniques are now also available for increasing the resolution of fluorescence microscopy to below 40 nm, commonly referred to as super-resolution microscopy, rendering the increased resolution of PLA compared to IF obsolete (234).

3.3 Assessing G-quadruplex stabilization

A protein or small molecule that binds a G-quadruplex but not ssDNA in an extended conformation is expected to stabilize the structure. Several methods exist for directly assessing this ability *in vitro*. The recent development of G-quadruplex specific antibodies has also made it possible to study this in cells (16,17).

3.3.1 Exonuclease I protection assay

In this assay G-quadruplex stability is measured by the ability of G-quadruplexes to resist exonucleolytic cleavage by the 3’→5’ Exonuclease I (Exo I) from *E. coli*. This assay was originally developed for assessing the G-quadruplex stabilizing effect of different buffers and small molecule ligands (235), but was adapted for use with protein in paper I of this thesis. 5’ labeled oligonucleotides are used, and the level of exonucleolytic cleavage is measured by running the products on a denaturing polyacrylamide gel. The reactions are performed both on oligonucleotides that forms a G-quadruplex at the 3’ end, and with a control oligonucleotide that has a similar sequence but is unable to form G-quadruplexes. The strength of this assay is its simplicity.

Several other assays for measuring G-quadruplex stabilization exist that were not used here. Another equally simple method is a polymerase stop assay, where one measures the inhibitory effect G-quadruplexes have on the progression of a DNA-polymerase, rather than an exonuclease (236,237). EMSA can also be used (as previously described). The downside of these assays compared to some more sophisticated methods is that you obtain no specific information about how the quadruplex is folded when protein is bound. Several techniques exist for answering this. An assay that is often used for this purpose is dimethyl sulfate (DMS) footprinting (219). The principle of this assay is that G-bases involved in Hoogsteen base pairing will be protected from methylation by DMS, and subsequent cleavage by piperidine. Protected bases are then identified by running a sequencing gel. Another approach is to use FRET. In this method an oligonucleotide carries a donor and acceptor fluorophore that will be in closer proximity when a quadruplex is formed, thus giving higher FRET efficiency (238). This allows real time measurements of single molecules (238).

3.4 RNAi mediated gene knock-down

A powerful method for studying the function of a gene of interest is to reduce its expression levels in cells. This can be done in a number of ways, but one of the most popular approaches is to use RNA-interference (RNAi). RNAi is part of the cells endogenous systems for posttranscriptional regulation of gene expression, by selectively cleaving mRNA (239). In the cell micro RNA (miRNA) can be expressed, processed, and incorporated into the RNA-induced Silencing Complex (RISC) which degrades mRNA (240). The processed miRNA, called small interfering RNA (siRNA), will be the antisense of the mRNA targeted by RISC and thus act as a guide. One way of

using the cells endogenous RNAi system to selectively reduce the expression of a gene of interest is by the use of short hairpin RNA (shRNA), as was done in paper I and II of this. shRNA mediated knock-down starts with cells being transfected (i.e. DNA is introduced into cells with the use of a chemical reagent) with constructs expressing the shRNA. The shRNA carries a guide sequence towards the gene of interest that will be incorporated into RISC after processing by the cells endogenous RNAi machinery. Expressing shRNA from a plasmid allows for stable knock-down and selection of transfected cells, by markers contained in the plasmid. Cells can also be transfected directly with siRNA (resembling the endogenous siRNA that is incorporated into RISC) towards the gene of interest, but this only allows for transient knock-down without selection.

3.5 Studying transcription

3.5.1 Global transcription assay

In paper II of this thesis global transcription was assayed, meaning the amount of all nascent transcripts was measured. This was done by feeding the cells a modified nucleoside, 5-ethynyl uridine (EU), by adding it to the cell culture media 1 h before fixation. EU contains an alkyne group which reacts with azides. The RNA can thus be fluorescently labeled by letting it react with an azide-containing fluorophore. After the RNA is labeled immunofluorescence of a protein of interest can also be performed in the same cells. Imaging is then done using fluorescence microscopy. This method has the advantage of being very robust and simple, while avoiding the use of radioactively labeled nucleosides.

3.5.2 Reverse transcription quantitative PCR

There are many techniques for determining gene expression by measuring the amount of specific transcripts. Microarray and sequencing techniques are powerful for determining the expression of many genes individually (or even all of them). In the work presented here however, the interest has only been in a few genes when assessing knock-down efficiencies in paper I and II. Reverse Transcription quantitative Polymerase Chain Reaction (RT-qPCR) is a powerful yet relatively simple method for determining gene expression of a few genes at a time. In this assay RNA is purified from cell extracts and is reverse transcribed into complementary DNA (cDNA). The cDNA is amplified and detected in real time (RT can also be read as real time) by qPCR with primers specific to the gene of interest. The amount of amplified DNA is measured by a dye in the reaction mixture that only fluoresces when

bound to dsDNA. The number of PCR-cycles before reaching a certain threshold is determined and compared between samples. The values are normalized to that of a housekeeping gene, i.e. a gene that should be similarly expressed in all samples. Another system (not used here) uses a gene specific probe containing a reporter and quencher fluorophore that are freed from each other when cleaved by the polymerase used for PCR. This has the advantage of not detecting primer dimers and other non-specific amplification products, so the assay relies less on carefully validating primers.

4 RESULTS

4.1 Paper I

hnRNP U/SAF-A is a G-quadruplex binding protein that associates with telomeres in a cell cycle dependent manner

Many general features of telomere function are conserved between species. On a molecular level, the ability to form G-quadruplexes is highly conserved between the telomeric DNA sequences of eukaryotes, even though the exact sequence varies. Still little is known about the purpose of this alternative DNA structure at telomeres.

Heterogenous ribonuleoprotein U (hnRNP U)/scaffold attachment factor A (SAF-A) has previously been shown to be part of a telomerase holoenzyme and to act as a negative regulator of telomere length, but its function in telomere maintenance has since remained unexplored (118). The fact that telomeres are scaffold attachment regions, and that hnRNP U has a glycine-rich RGG-box containing C-terminus, a common feature of G-quadruplex binding proteins, led us to further explore the role of hnRNP U in telomere maintenance.

Using PLA we found that hnRNP U is at close proximity to the telomeric protein TRF2 in mESCs and in human U2OS cells (that depend on ALT). ChIP confirmed that endogenous hnRNP U is attached to telomeric DNA in cells. Cell cycle synchronization followed by ChIP revealed that hnRNP U association peaks in G1/S to early S-phase, decreases at mid to late S-phase, then remains low throughout the rest of the cell cycle. Next, we explored how hnRNP U recognizes telomeric DNA. By using a DNA pull-down assay where biotinylated oligonucleotides are used to pull-down endogenous proteins from cell extracts, we show that hnRNP U preferentially binds a telomeric overhang that can form a G-quadruplex. Besides modifying the telomeric sequence by exchanging every other dG to 7-deaza-8-aza-dG, leaving it unable to form G-quadruplexes, we also evaluated the effect of G-quadruplex formation on hnRNP U binding by replacing K^+ with Li^+ in the buffer, and by adding increasing amounts of the G-quadruplex ligand TMPyP4, both of which confirmed the importance of G-quadruplex formation for hnRNP U binding. Given the known role of RGG-box motifs in G-quadruplex binding, we studied whether the C-terminus of hnRNP U is needed for G-quadruplex binding. GFP-tagged full length hnRNP U was pulled down by a telomeric overhang oligonucleotide, while hnRNP U with a

C-terminal truncation was not, demonstrating that the hnRNP U C-terminus (CT) is essential for G-quadruplex binding. To confirm that this was a direct interaction we repeated our DNA binding assays with recombinantly expressed and purified hnRNP U CT. While hnRNP U CT alone binds all the oligonucleotides tested, it binds G-quadruplexes with specificity, as evident upon the addition of non-specific competitor DNA. A non-biotinylated TERRA oligonucleotide is a poor competitor to both a single-stranded telomeric oligonucleotide and one that contains a 3' overhang, but it competes more strongly with the former than the latter. Hence, TERRA may compete more efficiently with hnRNP U binding at the 3' terminus than the ss/dsDNA border of the telomeric overhang *in vivo*.

In DNA pull-down assays of endogenous proteins we found that, as expected from previously published studies, replication protein A (RPA) binds to a lesser extent to oligonucleotides forming G-quadruplexes. Since RPA accumulation can lead to a DNA damage response, G-quadruplex formation has been proposed to help protect telomeres by excluding RPA. We next evaluated whether hnRNP U can aid in the exclusion of RPA from telomeres by stabilizing G-quadruplexes. To this end we first confirmed that hnRNP U CT stabilizes G-quadruplexes by using an exonuclease assay. Next, we tested whether pre-incubating a telomeric oligonucleotide with hnRNP U CT, to promote G-quadruplex formation, has an effect on RPA binding. Endogenous RPA in cell extracts showed diminished binding to a telomeric oligonuleotide pre-incubated with hnRNP U CT. However, RPA binding to a telomeric oligonucleotide carrying the deaza-substitution was not reduced following pre-incubation with hnRNP U CT, demonstrating that the effect is dependent on G-quadruplex formation. To evaluate if this mechanism may have any relevance *in vivo* we next knocked down *Hnrnpu* using shRNA (sh Hnrnpu) in mESCs and used IF to study RPA2 co-localization with TRF2. sh Hnrnpu treated cells showed an increase in the number of RPA2-foci co-localizing with TRF2-foci, suggesting that hnRNP U has a role in preventing telomere binding by RPA in cells.

4.2 Paper II

SAF-A forms a complex with BRG1 and both components are required for RNA polymerase II mediated transcription

SAF-A/hnRNP U and BRG1/SMARCA4 share several important characteristics. They are both found in the nuclear matrix (241,242), are essential during development (243,244), have been implicated in chromatin loop formation (241,245), and have important functions during transcription, evident by both of them interacting with Pol II (194,211). These common characteristics prompted us to investigate if they formed a complex, and if so its relevance for transcription.

By performing co-IP, IF and PLA it was established that SAF-A and BRG1 do form abundant complexes located throughout the nucleus of mESCs, except for nucleoli where both proteins were clearly absent. PLA following treatment with RNase A still showed abundant complexes, while RNase A treatment before co-IP reduced but did not abolish SAF-A associated BRG1. This indicates that the complex is not solely dependent on the presence of RNA. The complexes also remain abundant after differentiation of mESCs is induced by withdrawal of leukemia inhibitory factor (LIF) and addition of retinoic acid (RA). Next, the function of *Saf-a* and *Brg1* was evaluated by shRNA-mediated knock-down. After 96 h of knock-down the cells were in bad condition, and very little RNA could be isolated, which lead us to investigate a possible role of the two genes in global transcription. After 48 hours of knock-down only a modest decrease in 5-ethyl uridine incorporation into nascent RNA is seen for each gene individually, but a slightly more marked decrease is seen when both genes were knocked down simultaneously. After 72 hours of knock-down the decrease in nascent RNA is more marked, being roughly 50% for each gene individually, while simultaneous knock-down showed an almost complete loss of nascent RNA. Notably, nucleolar transcription was unaffected, which is expected given the individual proteins and the complex they form being absent from nucleoli. This demonstrates that both SAF-A and BRG1 are important for Pol II mediated transcription.

4.3 Paper III

Phosphorylated nucleolin interacts with translationally controlled tumor protein during mitosis and with Oct4 during interphase in ES cells

Translationally controlled tumor protein (TCTP), also called tumor protein translationally controlled 1 (Tpt1), is known to be important for reprogramming of cells through somatic cell nuclear transfer (246,247), and is important for regulating pluripotency (248). It has also been shown to be an important regulator of cell cycle progression (249). This led us to search for its interaction partners in ES cells.

Nucleolin (Ncl) was pulled down from mES cell extracts using recombinant Tpt1 coupled to cyanogen bromide activated sepharose beads. The complex was studied further using IF, PLA and co-IP. Little co-localization of Tpt1 and Ncl was seen in the nucleoli of mESCs. Instead, co-localization was most evident between phosphorylated nucleolin (Ncl-P) and Tpt1 in mitotic cells. The interaction of Tpt1 with Ncl-P was confirmed by co-IP on extracts from mitotically arrested mESCs. Likewise, PLA showed abundant complexes between Tpt1 and Ncl-P in mitotic cells, but very few complexes in interphase cells. The co-localization in mitotic cells decreased when cells were induced to differentiate by treatment with RA for 72 h.

The interaction between Ncl and Tpt1 made us wonder whether Ncl may have a role in maintaining pluripotency in mESCs. Since Ncl is known to be involved in transcriptional regulation we chose to look at its interaction with one of the key transcription factors for maintaining pluripotency; Oct4. IF revealed colocalization of Ncl-P with Oct4 in interphase nuclei of mESCs. The Oct4-Ncl-P complex in interphase nuclei was confirmed by PLA. Complexes with similar localization could also be shown using a phosphorylation unspecific Ncl antibody in the PLA assay. This set of antibodies were also used with PLA to show that the complex also exist in human ES cells.

5 DISCUSSION

In paper I we have shown that hnRNP U is a telomere binding protein that associates with telomeres in a cell cycle dependent manner, and that it interacts directly with telomeric G-quadruplexes via its C-terminus. We have also shown that hnRNP U may have a role in preventing RPA accumulation at telomeres, and our biochemical experiments suggest this involves G-quadruplex stabilization. This adds to the recent work of others that suggest that G-quadruplexes do form at telomeres and are important for telomere maintenance. We have only just begun to unravel the functions of hnRNP U at telomeres. Considering its effect on RPA binding to telomeric ssDNA, it is likely to influence also the binding of other proteins that bind ssDNA in an extended conformation. Judging by the cell cycle dependent binding of hnRNP U to telomeres, this might be important for regulating events prior to or early during replication of telomeres. Possibly it prevents events that occur later in S- phase and G2, such as 5' end resection (180), which is a known function of G-guadruplex stabilization at unprotected yeast telomeres (186). G-quadruplex stabilization by hnRNP U might also leave the 3' end inaccessible to telomerase. Given that hnRNP U is involved in transcriptional regulation and chromatin remodeling at other genomic loci, we can speculate that it may have a role also in regulating telomere heterochromatinization and TERRA expression. Since hnRNP U is a major scaffold attachment factor, it is likely to play a role in attaching the telomere to the matrix. Such spatio-temporal regulation might be important for the coordination of various processes at telomeres. An intriguing question that remains largely unexplored is whether telomeres attach to different parts of the nuclear matrix with specific functions. The fact that hnRNP A2* is also part of the nuclear matrix, resolves telomeric G-quadruplexes, and has the opposite effect on telomere length to hnRNP U, suggests that this might be the case (118,140).

In paper II we have shown that hnRNP U/SAF-A interacts with BRG1, and that both components of the complex are important for global transcription. The effect of knocking down both genes appears to be additive compared with knocking down the genes individually. This suggests that there may be some redundancy among the possible interaction partners for facilitating transcription. The fact that both proteins affect global transcription, together with the previously known roles of both proteins in transcription, strongly suggests that the complex also acts in the regulation of global transcription. This allows for speculation about how the complex may act in facilitating transcription. Both proteins are known to associate with the nuclear matrix and have been implicated in chromatin loop formation (241,242,245). Hence

it is possible that the complex is involved in organizing the chromatin in a transcriptionally active loop. Both proteins interact with Pol II (194,211). While BRG1 promotes transcriptional elongation (212), hnRNP U/SAF-A can prevent elongation by inhibiting CTD phosphorylation by TFIIH (194). Hence, they may act as a brake and throttle during elongation respectively.

In paper III we have shown the existence of two different protein-protein complexes in ESCs, each containing Ncl-P. The Tpt1-Ncl-P complex exists mainly in mitotic cells, while the Oct4-Ncl-P complex exists in the nuclei of interphase cells. The interaction with Tpt1 may be involved in regulation of cell cycle progression, in which both interaction partners are known to play significant roles (249-251). The interaction with Oct4 is likely to function in transcriptional regulation, considering both components are transcriptional regulators (252,253). Given that Ncl binds G-quadruplexes there is an interesting possibility that G-quadruplexes may play a part in regulating Oct4 driven gene expression (254). The octamere consensus sequence for Oct4 binding does not contain a PQS, but if a G-quadruplex forming sequence is located nearby Ncl might bind it. This could lead to either transcriptional repression or activation.

The papers presented in this thesis characterize novel protein-DNA and protein-protein interactions. Intriguingly, one component in each protein complex is a G-quadruplex binding protein. While it is still unknown if hnRNP U binds non-telomeric G-quadruplexes it is probable. It can thus be speculated that G-quadruplex binding by hnRNP U may also be important for its role in maintaining global transcription. Likewise, Ncl has an established role as a G-quadruplex binding transcription factor. G-quadruplexes might form more readily in ES cells, which may be important for maintaining the specific chromatin state that is imperative for pluripotency (255), or may be a consequence thereof. The G-quadruplex antibody raised and characterized by Lansdorp and colleagues was used on mES metaphase chromosomes, showing abundant staining (17). Whether differentiation of ES cells leads to a decrease in G-quadruplexes remains unclear. The possibility of G-quadruplexes having such a central role in key biological processes is exciting as this DNA structure, unlike many proteins, shows great promise for being targeted by small molecule ligands that may be used as drugs.

6 CONCLUSION

From paper I the following can be concluded:

- hnRNP U/SAF-A associates with telomeres.
- This association peaks during G1/S and early S-phase of the cell cycle in mESCs and is then decreased from mid S-phase and through the rest of the cell cycle.
- hnRNP U/SAF-A preferentially binds telomeric overhangs that form G-quadruplexes. The interaction is facilitated by its RGG-box and glycine rich C-terminus, which binds to and promotes the formation of telomeric G-quadruplexes.
- hnRNP U/SAF-A prevents RPA from binding telomeric ssDNA.

From paper II the following can be concluded:

- hnRNP U/SAF-A and BRG1 form a complex in mESCs.
- The complex is unaffected by early differentiation.
- hnRNP U/SAF-A and BRG1 are important for facilitating global transcription by Pol II.

From paper III the following can be concluded:

- Ncl-P interacts with Tpt1 in mESCs.
- The Tpt1-Ncl-P complex is found predominantly in mitotic cells.
- Tpt1 and Ncl-P interacts to a lesser extent when cells are induced to differentiate.
- Ncl-P interacts with Oct4 in ESCs.
- The Oct4-Ncl-P complex is found predominantly in the nuclei of interphase cells.
- These findings suggest that Ncl may have a role in regulating pluripotency.

7 FUTURE PERSPECTIVES

Much of the functional aspects of the protein-protein and protein-DNA interactions presented in this thesis remain unexplored.

To fully elucidate the role of the protein-protein interactions one will first need to clearly establish how the proteins interact, then study the effect of mutating the amino acids involved. For hnRNP U/SAF-A-BRG1 it would be interesting to further investigate which part or parts of the transcription process it is required for. For example, would a mutation in BRG1 that disrupts its interaction with hnRNP U affect the ability of BRG1 to promote transcriptional elongation? For the Ncl-containing complexes that were discovered, still nothing is known about their functionality beyond what can be speculated upon based on its subcellular localization and the previously known roles of the interaction partners. If the interactions can be disrupted, processes that need to be further investigated are cell cycle progression for Tpt1-Ncl-P and transcriptional regulation for Oct4-Ncl-P. Considering that the interactions take place mainly with phosphorylated Ncl it would also be relevant to further study which phosphorylations may be required for complex formation, and which kinases are involved.

The role of hnRNP U at telomeres also needs to be further characterized. Considering the previously known roles of G-quadruplexes in controlling telomerase activity (132,133), 5' end resection (186), and in regulating telomeric heterochromatin (160), these would be interesting topics to look further into. Also, it would be interesting to see if hnRNP U has a role in regulating TERRA transcription. Investigating these possible functions could also help explain how hnRNP U is involved in telomere length regulation.

Another interesting aspect to study further in relation to all the papers in this thesis is how G-quadruplexes may regulate function. In the case of hnRNP U a relevant question is whether it binds G-quadruplexes other than the telomeric. Considering that it binds preferentially to G-quadruplexes adjacent to dsDNA, it would be interesting to see if it also binds other alternative DNA structures such as G-quadruplex containing R-loops. If so, does this affect the function of the BRG1-hnRNP U complex in transcription? Similarily, Ncl is already known to bind promoter G-quadruplexes. How does this affect the function of the Oct4-Ncl-P complex? The use of G-quadruplex ligands could help answer this question. Potentially, an even more powerful tool in studying these aspects is the newly developed antibodies that can

recognize G-quadruplexes *in situ* and can be used for DNA immuno-precipitation.

Last but not least, combining the DNA pull-down technique presented here with G-quadruplex antibodies may provide powerful tools for making a larger study of all proteins that bind G-quadruplexes in cells and their possible functions. Rather than looking at individual proteins by Western blotting following DNA pull-down, one could run tandem mass spectrometry to identify all bound proteins and their relative quantities. By comparing the proteins bound to a G-quadruplex forming oligonucleotide with those bound to a non-quadruplex forming (i.e. deaza-substituted) oligonucleotide, one could identify all proteins that specifically bind the G-quadruplex. This could be done with several known G-quadruplex forming sequences, as well as exogenous ones. This could possibly distinguish between proteins that bind specifically to certain G-quadruplexes and those that show general G-quadruplex binding. Antibodies would then serve as excellent tools to validate that these proteins do indeed bind G-quadruplexes in cells, and genetic knock-down could then show if they promote G-quadruplex formation. Considering that many of the identified proteins are likely to already have known functions, one could possibly infer from this a lot of the biological functions that G-quadruplexes may have.

ACKNOWLEDGEMENT

Jag vill tacka **Keiko Funa** för att hon låtit mig arbeta i hennes labb och för att hon handlett mig trotts att jag haft ett projekt något utanför gruppens övriga intressesfär. Du är en person som verkligen bryr sig om andra människor och som sätter andras intressen framför sina egna.

Jag vill också tacka för all den hjälp och stöd jag fått från medlemmarna i de forskningsgrupper jag haft nöjet att arbeta i.

Dzeneta Vizlin-Hodzic är den enda som funnits där både i början och slutet. Utan din hjälp och stöd hade en svår situation kanske blivit omöjlig. Jag vill också ge ett särskilt tack till **Lydia Green** som hjälpt mig både med mitt arbete, och som kritiskt läst igenom denna avhandling. I would also like to thank all the present and former members of the Keiko Funa group, **Erik**, **Kristell**, **Praveen**, **Qiwei**, **Ravi** and **Tomiko** whom I have had the pleasure to work with. Even though I have worked with things quite different to the rest of you, I have always felt welcome as a member of the group.

Tack också till min biträdande handledare **Pierre Åman**. Min sejour i din grupp slutade inte riktigt som någon av oss tänkt sig eller hoppats, men det var roligt och intressant så länge det varade, och jag lärde mig mycket. Tack även till övriga gruppmedlemmar **Anders**, **Christer** och **Pernilla**, för all den hjälp jag fått.

Tack **Stina Simonsson** för att jag fick chansen att göra mitt examensarbete hos dig. Du lät mig prova på svåra och inspirerande försök redan från början, vilket både ökade mitt intresse för vetenskap och ingöt självförtroende. Tack **Tomas Simonsson** för att jag fick börja doktorera hos dig, och för att du introducerade mig till de spännande forskningsområden som präglar denna avhandling. Vi har haft många intressanta vetenskapliga diskussioner, vilka jag är tacksam för.

Jessica Ryme, tyvärr fick vi inte jobba ihop så länge, men jag uppskattade den tid som var. Du var den enda som fanns där en tid, och jag har insett att du och **Stefan** gav mig bra råd. Tack också till **Helena Johansson**. Det var en ganska kort tid även vi arbetade ihop, men det var också en tid då jag lärde mig mycket från er andra i labbet.

Tack **Per Elias** och **Anne Uv**, som har hjälpt och stöttat mig mycket på slutet av min doktorandtid.

I would also like to thank all of the people I have had the pleasure of getting to know while working at medical biochemistry, Shalgrenska cancer center, and when teaching.

Jag vill också tacka min släkt och familj som stöttat mig mycket. Speciellt min mor **Agneta**, som orkat lyssna på allt elände jag beklagat mig över de senaste åren, och som kommit med många goda råd som jag uppskattat mycket. Tack också för att jag alltid varit välkommen hem till Lund och Dalby. Det har varit skönt att kunna komma hem och koppla av då och då. Tack också min övriga släkt som förgyllt mina besök i Skåne. Framförallt **Johan**, **Petra**, och kanske mer än någon annan lille **Alexander**, som alltid är lika glad att se en. Det ska bli roligt även att lära känna din lillebror **Jonathan**. Tack också till släkten i Göteborg. Det har känts tryggt med er i närheten. Tack mormor **Siv**, för att du ställt upp på alla möjliga vis. Allt från att ha gett mig husrum till att storhandla på Ica. Tack **Chatrine och Chris** för alla söndagsmiddagar. Det har varit bra med lite variation från min annars ganska pastadominerade diet.

REFERENCES

1. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
2. Bang, I. (1910) Untersuchungen über die Guanylsäure. *Biochemische Zeitschrift*, **26**, 293-311.
3. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc Natl Acad Sci U S A*, **48**, 2013-2018.
4. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov*, **10**, 261-275.
5. Simonsson, T. (2001) G-quadruplex DNA structures--variations on a theme. *Biol Chem*, **382**, 621-628.
6. Arnott, S., Chandrasekaran, R. and Marttila, C.M. (1974) Structures for polyinosinic acid and polyguanylic acid. *Biochem J*, **141**, 537-543.
7. Zimmerman, S.B., Cohen, G.H. and Davies, D.R. (1975) X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J Mol Biol*, **92**, 181-192.
8. Gray, D.M. (1974) A circular dichroism study of poly dG, poly dC, and poly dG:dC. *Biopolymers*, **13**, 2087-2102.
9. Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364-366.
10. Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I., Jr. and Blackburn, E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. *Cell*, **51**, 899-908.
11. Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825-829.
12. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell*, **59**, 871-880.
13. Simonsson, T., Kubista, M. and Pecinka, P. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Research*, **26**, 1167-1172.
14. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Research*, **33**, 2908-2916.
15. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research*, **35**, 406-413.

16. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem*, **5**, 182-186.
17. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Sen, D. and Lansdorp, P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Research*, **42**, 860-869.
18. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell*, **59**, 871-880.
19. Sen, D. and Gilbert, W. (1992) Guanine quartet structures. *Methods Enzymol*, **211**, 191-199.
20. Phan, A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS Journal*, **277**, 1107-1117.
21. Izbicka, E., Wheelhouse, R.T., Raymond, E., Davidson, K.K., Lawrence, R.A., Sun, D., Windle, B.E., Hurley, L.H. and Von Hoff, D.D. (1999) Effects of Cationic Porphyrins as G-Quadruplex Interactive Agents in Human Tumor Cells. *Cancer Research*, **59**, 639-644.
22. Parkinson, G.N., Ghosh, R. and Neidle, S. (2007) Structural Basis for Binding of Porphyrin to Human Telomeres. *Biochemistry*, **46**, 2390-2397.
23. Evans, M.J. and Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature*, **292**, 154-156.
24. Takahashi, K. and Yamanaka, S. (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, **126**, 663-676.
25. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, **131**, 861-872.
26. Arellano, M. and Moreno, S. (1997) Regulation of CDK/cyclin complexes during the cell cycle. *The International Journal of Biochemistry & Cell Biology*, **29**, 559-573.
27. Savatier, P., Huang, S., Szekely, L., Wiman, K.G. and Samarut, J. (1994) Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene*, **9**, 809-818.
28. Stead, E., White, J., Faast, R., Conn, S., Goldstone, S., Rathjen, J., Dhingra, U., Rathjen, P., Walker, D. and Dalton, S. (2002) Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene*, **21**, 8320-8333.
29. Blackburn, E.H. and Gall, J.G. (1978) A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *Journal of Molecular Biology*, **120**, 33-53.

30. Lejnine, S., Makarov, V.L. and Langmore, J.P. (1995) Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. *Proceedings of the National Academy of Sciences*, **92**, 2393-2397.
31. de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. and Varmus, H.E. (1990) Structure and variability of human chromosome ends. *Molecular and Cellular Biology*, **10**, 518-527.
32. Kipling, D. and Cooke, H.J. (1990) Hypervariable ultra-long telomeres in mice. *Nature*, **347**, 400-402.
33. McElligott, R. and Wellinger, R.J. (1997) *The terminal DNA structure of mammalian chromosomes*.
34. Denchi, E.L. and de Lange, T. (2007) Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature*, **448**, 1068-1071.
35. Takai, H., Smogorzewska, A. and de Lange, T. (2003) DNA Damage Foci at Dysfunctional Telomeres. *Current Biology*, **13**, 1549-1556.
36. Karlseder, J., Broccoli, D., Dai, Y., Hardy, S. and de Lange, T. (1999) p53- and ATM-Dependent Apoptosis Induced by Telomeres Lacking TRF2. *Science*, **283**, 1321-1325.
37. Zou, L. and Elledge, S.J. (2003) Sensing DNA Damage Through ATRIP Recognition of RPA-ssDNA Complexes. *Science*, **300**, 1542-1548.
38. Blasina, A., de Weyer, I.V., Laus, M.C., Luyten, W.H.M.L., Parker, A.E. and McGowan, C.H. (1999) A human homologue of the checkpoint kinase Cds1 directly inhibits Cdc25 phosphatase. *Current Biology*, **9**, 1-10.
39. Fagagna, F.d.A.d., Reaper, P.M., Clay-Farrace, L., Fiegler, H., Carr, P., von Zglinicki, T., Saretzki, G., Carter, N.P. and Jackson, S.P. (2003) A DNA damage checkpoint response in telomere-initiated senescence. *Nature*, **426**, 194-198.
40. Aladjem, M.I., Spike, B.T., Rodewald, L.W., Hope, T.J., Klemm, M., Jaenisch, R. and Wahl, G.M. (1998) ES cells do not activate p53-dependent stress responses and undergo p53-independent apoptosis in response to DNA damage. *Current Biology*, **8**, 145-155.
41. Hong, Y. and Stambrook, P.J. (2004) Restoration of an absent G1 arrest and protection from apoptosis in embryonic stem cells after ionizing radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14443-14448.
42. Liu, Q., Guntuku, S., Cui, X.-S., Matsuoka, S., Cortez, D., Tamai, K., Luo, G., Carattini-Rivera, S., DeMayo, F., Bradley, A. *et al.* (2000) Chk1 is an essential kinase that is regulated by Atr and required for the G2/M DNA damage checkpoint. *Genes & Development*, **14**, 1448-1459.

43. Corbet, S.W., Clarke, A.R., Gledhill, S. and Wyllie, A.H. (1999) P53-dependent and -independent links between DNA-damage, apoptosis and mutation frequency in ES cells. *Oncogene*, **18**, 1537-1544.
44. Lin, T., Chao, C., Saito, S.i., Mazur, S.J., Murphy, M.E., Appella, E. and Xu, Y. (2005) p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat Cell Biol*, **7**, 165-171.
45. Cervantes, R.B., Stringer, J.R., Shao, C., Tischfield, J.A. and Stambrook, P.J. (2002) Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proceedings of the National Academy of Sciences*, **99**, 3586-3590.
46. Schneider, R.P., Garrobo, I., Foronda, M., Palacios, J.A., Marión, R.M., Flores, I., Ortega, S. and Blasco, M.A. (2013) TRF1 is a stem cell marker and is essential for the generation of induced pluripotent stem cells. *Nat Commun*, **4**.
47. Tejera, A.M., Stagno d'Alcontres, M., Thanasoula, M., Marion, R.M., Martinez, P., Liao, C., Flores, J.M., Tarsounas, M. and Blasco, M.A. (2010) TPP1 Is Required for TERT Recruitment, Telomere Elongation during Nuclear Reprogramming, and Normal Skin Development in Mice. *Developmental Cell*, **18**, 775-789.
48. Sfeir, A. and de Lange, T. (2012) Removal of Shelterin Reveals the Telomere End-Protection Problem. *Science*, **336**, 593-597.
49. Palm, W. and de Lange, T. (2008) How Shelterin Protects Mammalian Telomeres. *Annual Review of Genetics*, **42**, 301-334.
50. Broccoli, D., Smogorzewska, A., Chong, L. and de Lange, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat Genet*, **17**, 231-235.
51. Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P. and de Lange, T. (1995) A human telomeric protein. *Science*, **270**, 1663-1667.
52. Baumann, P. and Cech, T.R. (2001) Pot1, the Putative Telomere End-Binding Protein in Fission Yeast and Humans. *Science*, **292**, 1171-1175.
53. Kim, S.-h., Kaminker, P. and Campisi, J. (1999) TIN2, a new regulator of telomere length in human cells. *Nat Genet*, **23**, 405-412.
54. Houghtaling, B.R., Cuttonaro, L., Chang, W. and Smith, S. (2004) A Dynamic Molecular Link between the Telomere Length Regulator TRF1 and the Chromosome End Protector TRF2. *Current Biology*, **14**, 1621-1631.
55. Li, B., Oestreich, S. and de Lange, T. (2000) Identification of Human Rap1: Implications for Telomere Evolution. *Cell*, **101**, 471-483.
56. van Steensel, B., Smogorzewska, A. and de Lange, T. (1998) TRF2 Protects Human Telomeres from End-to-End Fusions. *Cell*, **92**, 401-413.

57. Gong, Y. and de Lange, T. (2010) A Shld1-Controlled POT1a Provides Support for Repression of ATR Signaling at Telomeres through RPA Exclusion. *Molecular Cell*, **40**, 377-387.
58. Verdun, R.E., Crabbe, L., Haggblom, C. and Karlseder, J. (2005) Functional Human Telomeres Are Recognized as DNA Damage in G2 of the Cell Cycle. *Molecular Cell*, **20**, 551-561.
59. Verdun, R.E. and Karlseder, J. (2006) The DNA Damage Machinery and Homologous Recombination Pathway Act Consecutively to Protect Human Telomeres. *Cell*, **127**, 709-720.
60. Crabbe, L., Verdun, R.E., Haggblom, C.I. and Karlseder, J. (2004) Defective Telomere Lagging Strand Synthesis in Cells Lacking WRN Helicase Activity. *Science*, **306**, 1951-1953.
61. Barefield, C. and Karlseder, J. (2012) The BLM helicase contributes to telomere maintenance through processing of late-replicating intermediate structures. *Nucleic Acids Research*, **40**, 7358-7367.
62. Vannier, J.-B., Pavicic-Kaltenbrunner, V., Petalcorin, Mark I.R., Ding, H. and Boulton, Simon J. (2012) RTEL1 Dismantles T Loops and Counteracts Telomeric G4-DNA to Maintain Telomere Integrity. *Cell*, **149**, 795-806.
63. Ghosh, A.K., Rossi, M.L., Singh, D.K., Dunn, C., Ramamoorthy, M., Croteau, D.L., Liu, Y. and Bohr, V.A. (2012) RECQL4, the Protein Mutated in Rothmund-Thomson Syndrome, Functions in Telomere Maintenance. *Journal of Biological Chemistry*, **287**, 196-209.
64. Flynn, R.L., Centore, R.C., O'Sullivan, R.J., Rai, R., Tse, A., Songyang, Z., Chang, S., Karlseder, J. and Zou, L. (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature*, **471**, 532-536.
65. Schoeftner, S. and Blasco, M.A. (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol*, **10**, 228-236.
66. Porro, A., Feuerhahn, S., Reichenbach, P. and Lingner, J. (2010) Molecular Dissection of Telomeric Repeat-Containing RNA Biogenesis Unveils the Presence of Distinct and Multiple Regulatory Pathways. *Molecular and Cellular Biology*, **30**, 4808-4817.
67. Stansel, R.M., de Lange, T. and Griffith, J.D. (2001) T-loop assembly in vitro involves binding of TRF2 near the 3[prime] telomeric overhang. *EMBO J*, **20**, 5532-5540.
68. Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H. and de Lange, T. (1999) Mammalian Telomeres End in a Large Duplex Loop. *Cell*, **97**, 503-514.
69. Doksani, Y., Wu, John Y., de Lange, T. and Zhuang, X. (2013) Super-Resolution Fluorescence Imaging of Telomeres Reveals TRF2-Dependent T-loop Formation. *Cell*, **155**, 345-356.
70. Sfeir, A., Kosiyatrakul, S.T., Hockemeyer, D., MacRae, S.L., Karlseder, J., Schildkraut, C.L. and de Lange, T. (2009) Mammalian

Telomeres Resemble Fragile Sites and Require TRF1 for Efficient Replication. *Cell*, **138**, 90-103.
71. Ray, S., Bandaria, J.N., Qureshi, M.H., Yildiz, A. and Balci, H. (2014) G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proceedings of the National Academy of Sciences*, **111**, 2990-2995.
72. Cai, L., Roginskaya, M., Qu, Y., Yang, Z., Xu, Y. and Zou, Y. (2007) Structural Characterization of Human RPA Sequential Binding to Single-Stranded DNA Using ssDNA as a Molecular Ruler. *Biochemistry*, **46**, 8226-8233.
73. Lei, M., Podell, E.R. and Cech, T.R. (2004) Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat Struct Mol Biol*, **11**, 1223-1229.
74. Hwang, H., Buncher, N., Opresko, Patricia L. and Myong, S. (2012) POT1-TPP1 Regulates Telomeric Overhang Structural Dynamics. *Structure*, **20**, 1872-1880.
75. Salas, T.R., Petruseva, I., Lavrik, O., Bourdoncle, A., Mergny, J.-L., Favre, A. and Saintomé, C. (2006) Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Research*, **34**, 4857-4865.
76. Gomez, D., O'Donohue, M.-F., Wenner, T., Douarre, C., Macadré, J., Koebel, P., Giraud-Panis, M.-J., Kaplan, H., Kolkes, A., Shin-ya, K. *et al.* (2006) The G-quadruplex Ligand Telomestatin Inhibits POT1 Binding to Telomeric Sequences In vitro and Induces GFP-POT1 Dissociation from Telomeres in Human Cells. *Cancer Research*, **66**, 6908-6912.
77. Safa, L., Delagoutte, E., Petruseva, I., Alberti, P., Lavrik, O., Riou, J.-F. and Saintomé, C. Binding polarity of RPA to telomeric sequences and influence of G-quadruplex stability. *Biochimie*.
78. Salas, T.R., Petruseva, I., Lavrik, O. and Saintomé, C. (2009) Evidence for direct contact between the RPA3 subunit of the human replication protein A and single-stranded DNA. *Nucleic Acids Research*, **37**, 38-46.
79. Watson, J.D. (1972) Origin of concatemeric T7 DNA. *Nat New Biol*, **239**, 197-201.
80. Hayflick, L. and Moorhead, P.S. (1961) The serial cultivation of human diploid cell strains. *Experimental Cell Research*, **25**, 585-621.
81. Bryan, T.M., Englezou, A., Dalla-Pozza, L., Dunham, M.A. and Reddel, R.R. (1997) Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat Med*, **3**, 1271-1274.
82. Lundblad, V. and Blackburn, E.H. (1993) An alternative pathway for yeast telomere maintenance rescues est1− senescence. *Cell*, **73**, 347-360.

83. Liu, L., Bailey, S.M., Okuka, M., Munoz, P., Li, C., Zhou, L., Wu, C., Czerwiec, E., Sandler, L., Seyfang, A. *et al.* (2007) Telomere lengthening early in development. *Nat Cell Biol*, **9**, 1436-1441.
84. Lundblad, V. and Szostak, J.W. (1989) A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell*, **57**, 633-643.
85. Harley, C.B., Futcher, A.B. and Greider, C.W. (1990) Telomeres shorten during ageing of human fibroblasts. *Nature*, **345**, 458-460.
86. Vaziri, H. and Benchimol, S. (1998) Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. *Current Biology*, **8**, 279-282.
87. Herbig, U., Jobling, W.A., Chen, B.P.C., Chen, D.J. and Sedivy, J.M. (2004) Telomere Shortening Triggers Senescence of Human Cells through a Pathway Involving ATM, p53, and p21CIP1, but Not p16INK4a. *Molecular Cell*, **14**, 501-513.
88. Prowse, K.R. and Greider, C.W. (1995) Developmental and tissue-specific regulation of mouse telomerase and telomere length. *Proceedings of the National Academy of Sciences*, **92**, 4818-4822.
89. Hastie, N.D., Dempster, M., Dunlop, M.G., Thompson, A.M., Green, D.K. and Allshire, R.C. (1990) Telomere reduction in human colorectal carcinoma and with ageing. *Nature*, **346**, 866-868.
90. Mitchell, J.R., Wood, E. and Collins, K. (1999) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature*, **402**, 551-555.
91. Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P.J. and Dokal, I. (2001) The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature*, **413**, 432-435.
92. Blasco, M.A., Lee, H.-W., Hande, M.P., Samper, E., Lansdorp, P.M., DePinho, R.A. and Greider, C.W. (1997) Telomere Shortening and Tumor Formation by Mouse Cells Lacking Telomerase RNA. *Cell*, **91**, 25-34.
93. Rudolph, K.L., Chang, S., Lee, H.-W., Blasco, M., Gottlieb, G.J., Greider, C. and DePinho, R.A. (1999) Longevity, Stress Response, and Cancer in Aging Telomerase-Deficient Mice. *Cell*, **96**, 701-712.
94. Lee, H.-W., Blasco, M.A., Gottlieb, G.J., Horner, J.W., Greider, C.W. and DePinho, R.A. (1998) Essential role of mouse telomerase in highly proliferative organs. *Nature*, **392**, 569-574.
95. Cotsarelis, G., Sun, T.-T. and Lavker, R.M. (1990) Label-retaining cells reside in the bulge area of pilosebaceous unit: Implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell*, **61**, 1329-1337.
96. Taylor, G., Lehrer, M.S., Jensen, P.J., Sun, T.-T. and Lavker, R.M. (2000) Involvement of Follicular Stem Cells in Forming Not Only the Follicle but Also the Epidermis. *Cell*, **102**, 451-461.

97. Flores, I., Canela, A., Vera, E., Tejera, A., Cotsarelis, G. and Blasco, M.A. (2008) The longest telomeres: a general signature of adult stem cell compartments. *Genes & Development*, **22**, 654-667.
98. Niida, H., Matsumoto, T., Satoh, H., Shiwa, M., Tokutake, Y., Furuichi, Y. and Shinkai, Y. (1998) Severe growth defect in mouse cells lacking the telomerase RNA component. *Nat Genet*, **19**, 203-206.
99. Zalzman, M., Falco, G., Sharova, L.V., Nishiyama, A., Thomas, M., Lee, S.-L., Stagg, C.A., Hoang, H.G., Yang, H.-T., Indig, F.E. *et al.* (2010) Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature*, **464**, 858-863.
100. Marion, R.M., Strati, K., Li, H., Tejera, A., Schoeftner, S., Ortega, S., Serrano, M. and Blasco, M.A. (2009) Telomeres Acquire Embryonic Stem Cell Characteristics in Induced Pluripotent Stem Cells. *Cell Stem Cell*, **4**, 141-154.
101. Jiang, J., Lv, W., Ye, X., Wang, L., Zhang, M., Yang, H., Okuka, M., Zhou, C., Zhang, X., Liu, L. *et al.* (2013) Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells as demonstrated by tetraploid complementation. *Cell Res*, **23**, 92-106.
102. Shay, J.W. and Bacchetti, S. (1997) A survey of telomerase activity in human cancer. *European Journal of Cancer*, **33**, 787-791.
103. Heaphy, C.M., de Wilde, R.F., Jiao, Y., Klein, A.P., Edil, B.H., Shi, C., Bettegowda, C., Rodriguez, F.J., Eberhart, C.G., Hebbar, S. *et al.* (2011) Altered Telomeres in Tumors with ATRX and DAXX Mutations. *Science*, **333**, 425.
104. Henson, J.D., Hannay, J.A., McCarthy, S.W., Royds, J.A., Yeager, T.R., Robinson, R.A., Wharton, S.B., Jellinek, D.A., Arbuckle, S.M., Yoo, J. *et al.* (2005) A Robust Assay for Alternative Lengthening of Telomeres in Tumors Shows the Significance of Alternative Lengthening of Telomeres in Sarcomas and Astrocytomas. *Clinical Cancer Research*, **11**, 217-225.
105. Gonzalez-Suarez, E., Samper, E., Flores, J.M. and Blasco, M.A. (2000) Telomerase-deficient mice with short telomeres are resistant to skin tumorigenesis. *Nat Genet*, **26**, 114-117.
106. Gonzalez-Suarez, E., Samper, E., Ramirez, A., Flores, J.M., Martin-Caballero, J., Jorcano, J.L. and Blasco, M.A. (2001) Increased epidermal tumors and increased skin wound healing in transgenic mice overexpressing the catalytic subunit of telomerase, mTERT, in basal keratinocytes. *EMBO J*, **20**, 2619-2630.
107. Gonzalez-Suarez, E., Geserick, C., Flores, J.M. and Blasco, M.A. (2005) Antagonistic effects of telomerase on cancer and aging in K5-mTert transgenic mice. **24**, 2256-2270.
108. Tomás-Loba, A., Flores, I., Fernández-Marcos, P.J., Cayuela, M.L., Maraver, A., Tejera, A., Borrás, C., Matheu, A., Klatt, P., Flores,

J.M. *et al.* (2008) Telomerase Reverse Transcriptase Delays Aging in Cancer-Resistant Mice. *Cell*, **135**, 609-622.

109. Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B. and Cech, T.R. (1997) Telomerase Catalytic Subunit Homologs from Fission Yeast and Human. *Science*, **277**, 955-959.
110. Blasco, M., Funk, W., Villeponteau, B. and Greider, C. (1995) Functional characterization and developmental regulation of mouse telomerase RNA. *Science*, **269**, 1267-1270.
111. Gregg B, M. (1989) The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell*, **59**, 521-529.
112. Sauerwald, A., Sandin, S., Cristofari, G., Scheres, S.H.W., Lingner, J. and Rhodes, D. (2013) Structure of active dimeric human telomerase. *Nat Struct Mol Biol*, **20**, 454-460.
113. Xi, L. and Cech, T.R. (2014) Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT. *Nucleic Acids Research*, **42**, 8565-8577.
114. Ten Hagen, K.G., Gilbert, D.M., Willard, H.F. and Cohen, S.N. (1990) Replication timing of DNA sequences associated with human centromeres and telomeres. *Molecular and Cellular Biology*, **10**, 6348-6355.
115. Tomlinson, R.L., Ziegler, T.D., Supakorndej, T., Terns, R.M. and Terns, M.P. (2006) Cell Cycle-regulated Trafficking of Human Telomerase to Telomeres. *Molecular Biology of the Cell*, **17**, 955-965.
116. Marcand, S., Brevet, V. and Gilson, E. (1999) Progressive cis-inhibition of telomerase upon telomere elongation. *EMBO J*, **18**, 3509-3519.
117. Liu, Y., Kha, H., Ungrin, M., Robinson, M.O. and Harrington, L. (2002) Preferential maintenance of critically short telomeres in mammalian cells heterozygous for mTert. *Proceedings of the National Academy of Sciences*, **99**, 3597-3602.
118. Fu, D. and Collins, K. (2007) Purification of Human Telomerase Complexes Identifies Factors Involved in Telomerase Biogenesis and Telomere Length Regulation. *Molecular Cell*, **28**, 773-785.
119. Cohen, S.B., Graham, M.E., Lovrecz, G.O., Bache, N., Robinson, P.J. and Reddel, R.R. (2007) Protein Composition of Catalytically Active Human Telomerase from Immortal Cells. *Science*, **315**, 1850-1853.
120. Venteicher, A.S., Meng, Z., Mason, P.J., Veenstra, T.D. and Artandi, S.E. (2008) Identification of ATPases Pontin and Reptin as Telomerase Components Essential for Holoenzyme Assembly. *Cell*, **132**, 945-957.

121. Venteicher, A.S., Abreu, E.B., Meng, Z., McCann, K.E., Terns, R.M., Veenstra, T.D., Terns, M.P. and Artandi, S.E. (2009) A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science*, **323**, 644-648.
122. Stern, J.L., Zyner, K.G., Pickett, H.A., Cohen, S.B. and Bryan, T.M. (2012) Telomerase Recruitment Requires both TCAB1 and Cajal Bodies Independently. *Molecular and Cellular Biology*, **32**, 2384-2395.
123. Nandakumar, J., Bell, C.F., Weidenfeld, I., Zaug, A.J., Leinwand, L.A. and Cech, T.R. (2012) The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature*, **492**, 285-289.
124. Abreu, E., Aritonovska, E., Reichenbach, P., Cristofari, G., Culp, B., Terns, R.M., Lingner, J. and Terns, M.P. (2010) TIN2-Tethered TPP1 Recruits Human Telomerase to Telomeres In Vivo. *Molecular and Cellular Biology*, **30**, 2971-2982.
125. LaBranche, H., Dupuis, S., Ben-David, Y., Bani, M.-R., Wellinger, R.J. and Chabot, B. (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat Genet*, **19**, 199-202.
126. Zhang, Q.-S., Manche, L., Xu, R.-M. and Krainer, A.R. (2006) hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA*, **12**, 1116-1128.
127. Nagata, T., Takada, Y., Ono, A., Nagata, K., Konishi, Y., Nukina, T., Ono, M., Matsugami, A., Furukawa, A., Fujimoto, N. *et al.* (2008) Elucidation of the mode of interaction in the UP1-telomerase RNA-telomeric DNA ternary complex which serves to recruit telomerase to telomeric DNA and to enhance the telomerase activity. *Nucleic Acids Research*, **36**, 6816-6824.
128. Redon, S., Zemp, I. and Lingner, J. (2013) A three-state model for the regulation of telomerase by TERRA and hnRNPA1. *Nucleic Acids Research*, **41**, 9117-9128.
129. Zhao, Y., Abreu, E., Kim, J., Stadler, G., Eskiocak, U., Terns, Michael P., Terns, Rebecca M., Shay, Jerry W. and Wright, Woodring E. (2011) Processive and Distributive Extension of Human Telomeres by Telomerase under Homeostatic and Nonequilibrium Conditions. *Molecular Cell*, **42**, 297-307.
130. Smogorzewska, A., van Steensel, B., Bianchi, A., Oelmann, S., Schaefer, M.R., Schnapp, G. and de Lange, T. (2000) Control of Human Telomere Length by TRF1 and TRF2. *Molecular and Cellular Biology*, **20**, 1659-1668.
131. Kim, N.W. and Wu, F. (1997) Advances in quantification and characterization of telomerase activity by the telomeric repeat amplification protocol (TRAP). *Nucleic Acids Research*, **25**, 2595-2597.

132. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718-720.
133. Fletcher, T.M., Sun, D., Salazar, M. and Hurley, L.H. (1998) Effect of DNA Secondary Structure on Human Telomerase Activity. *Biochemistry*, **37**, 5536-5541.
134. De Cian, A., Cristofari, G., Reichenbach, P., De Lemos, E., Monchaud, D., Teulade-Fichou, M.-P., Shin-ya, K., Lacroix, L., Lingner, J. and Mergny, J.-L. (2007) Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proceedings of the National Academy of Sciences*, **104**, 17347-17352.
135. Xue, Y., Kan, Z.-y., Wang, Q., Yao, Y., Liu, J., Hao, Y.-h. and Tan, Z. (2007) Human Telomeric DNA Forms Parallel-Stranded Intramolecular G-Quadruplex in K+ Solution under Molecular Crowding Condition. *Journal of the American Chemical Society*, **129**, 11185-11191.
136. Oganesian, L., Moon, I.K., Bryan, T.M. and Jarstfer, M.B. (2006) Extension of G-quadruplex DNA by ciliate telomerase. *The EMBO Journal*, **25**, 1148-1159.
137. Zhang, M.L., Tong, X.J., Fu, X.H., Zhou, B.O., Wang, J., Liao, X.H., Li, Q.J., Shen, N., Ding, J. and Zhou, J.Q. (2010) Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation. *Nat Struct Mol Biol*, **17**, 202-209.
138. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. and Lipps, H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol*, **12**, 847-854.
139. Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D. and Lipps, H.J. (2008) Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat Struct Mol Biol*, **15**, 598-604.
140. Wang, F., Tang, M.L., Zeng, Z.X., Wu, R.Y., Xue, Y., Hao, Y.H., Pang, D.W., Zhao, Y. and Tan, Z. (2012) Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 20413-20418.
141. Wang, F., Podell, E.R., Zaug, A.J., Yang, Y., Baciu, P., Cech, T.R. and Lei, M. (2007) The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature*, **445**, 506-510.
142. Latrick, C.M. and Cech, T.R. (2010) POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *EMBO J*, **29**, 924-933.

143. Riou, J.F., Guittat, L., Mailliet, P., Laoui, A., Renou, E., Petitgenet, O., Mégnin-Chanet, F., Hélène, C. and Mergny, J.L. (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proceedings of the National Academy of Sciences*, **99**, 2672-2677.
144. Zhou, J.M., Zhu, X.F., Lu, Y.J., Deng, R., Huang, Z.S., Mei, Y.P., Wang, Y., Huang, W.L., Liu, Z.C., Gu, L.Q. *et al.* (2005) Senescence and telomere shortening induced by novel potent G-quadruplex interactive agents, quindoline derivatives, in human cancer cell lines. *Oncogene*, **25**, 503-511.
145. Tauchi, T., Shin-ya, K., Sashida, G., Sumi, M., Nakajima, A., Shimamoto, T., Ohyashiki, J.H. and Ohyashiki, K. (2003) Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095), against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene*, **22**, 5338-5347.
146. Tahara, H., Shin-ya, K., Seimiya, H., Yamada, H., Tsuruo, T. and Ide, T. (2005) G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3[prime] telomeric overhang in cancer cells. *Oncogene*, **25**, 1955-1966.
147. Garcia-Cao, M., O'Sullivan, R., Peters, A.H.F.M., Jenuwein, T. and Blasco, M.A. (2004) Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat Genet*, **36**, 94-99.
148. Gonzalo, S., Garcia-Cao, M., Fraga, M.F., Schotta, G., Peters, A.H.F.M., Cotter, S.E., Eguia, R., Dean, D.C., Esteller, M., Jenuwein, T. *et al.* (2005) Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. *Nat Cell Biol*, **7**, 420-428.
149. Benetti, R., Garcia-Cao, M. and Blasco, M.A. (2007) Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat Genet*, **39**, 243-250.
150. Benetti, R., Gonzalo, S., Jaco, I., Schotta, G., Klatt, P., Jenuwein, T. and Blasco, M.A. (2007) Suv4-20h deficiency results in telomere elongation and derepression of telomere recombination. *The Journal of Cell Biology*, **178**, 925-936.
151. Hazelrigg, T., Levis, R. and Rubin, G.M. (1984) Transformation of white locus DNA in Drosophila: Dosage compensation, zeste interaction, and position effects. *Cell*, **36**, 469-481.
152. Gottschling, D.E., Aparicio, O.M., Billington, B.L. and Zakian, V.A. (1990) Position effect at S. cerevisiae telomeres: Reversible repression of Pol II transcription. *Cell*, **63**, 751-762.
153. Stadler, G., Rahimov, F., King, O.D., Chen, J.C.J., Robin, J.D., Wagner, K.R., Shay, J.W., Emerson, C.P. and Wright, W.E. (2013)

Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat Struct Mol Biol*, **20**, 671-678.

154. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007) Telomeric Repeat–Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science*, **318**, 798-801.
155. Arnoult, N., Van Beneden, A. and Decottignies, A. (2012) Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1α. *Nat Struct Mol Biol*, **19**, 948-956.
156. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. and Lieberman, P.M. (2009) TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Molecular Cell*, **35**, 403-413.
157. Redon, S., Reichenbach, P. and Lingner, J. (2010) The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Research*, **38**, 5797-5806.
158. Biffi, G., Tannahill, D. and Balasubramanian, S. (2012) An Intramolecular G-Quadruplex Structure Is Required for Binding of Telomeric Repeat-Containing RNA to the Telomeric Protein TRF2. *Journal of the American Chemical Society*, **134**, 11974-11976.
159. Deng, Z., Dheekollu, J., Broccoli, D., Dutta, A. and Lieberman, P.M. (2007) The Origin Recognition Complex Localizes to Telomere Repeats and Prevents Telomere-Circle Formation. *Current Biology*, **17**, 1989-1995.
160. Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R. and Oyoshi, T. (2013) Regulation of Telomere Length by G-Quadruplex Telomere DNA- and TERRA-Binding Protein TLS/FUS. *Chemistry & Biology*, **20**, 341-350.
161. Pfeiffer, V., Crittin, J., Grolimund, L. and Lingner, J. (2013) The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J*, **32**, 2861-2871.
162. Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T. and Azzalin, C.M. (2014) RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun*, **5**.
163. Britton, S., Dernoncourt, E., Delteil, C., Froment, C., Schiltz, O., Salles, B., Frit, P. and Calsou, P. (2014) DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. *Nucleic Acids Research*, **42**, 9047-9062.
164. Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R. and Choo, K.H.A. (2010) ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Research*, **20**, 351-360.

165. Law, M.J., Lower, K.M., Voon, H.P.J., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A. *et al.* (2010) ATR-X Syndrome Protein Targets Tandem Repeats and Influences Allele-Specific Expression in a Size-Dependent Manner. *Cell*, **143**, 367-378.
166. Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X. *et al.* (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*, **140**, 678-691.
167. Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L. and Gasser, S.M. (1993) SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell*, **75**, 543-555.
168. Crabbe, L., Cesare, Anthony J., Kasuboski, James M., Fitzpatrick, James A.J. and Karlseder, J. (2012) Human Telomeres Are Tethered to the Nuclear Envelope during Postmitotic Nuclear Assembly. *Cell Reports*, **2**, 1521-1529.
169. de Lange, T. (1992) Human telomeres are attached to the nuclear matrix. *EMBO J*, **11**, 717-724.
170. Ludérus, M.E., van Steensel, B., Chong, L., Sibon, O.C., Cremers, F.F. and de Lange, T. (1996) Structure, subnuclear distribution, and nuclear matrix association of the mammalian telomeric complex. *The Journal of Cell Biology*, **135**, 867-881.
171. Kaminker, P.G., Kim, S.H., Desprez, P.Y. and Campisi, J. (2009) A novel form of the telomere-associated protein TIN2 localizes to the nuclear matrix. *Cell Cycle*, **8**, 931-939.
172. Gonzalez-Suarez, I., Redwood, A.B., Perkins, S.M., Vermolen, B., Lichtensztejin, D., Grotsky, D.A., Morgado-Palacin, L., Gapud, E.J., Sleckman, B.P., Sullivan, T. *et al.* (2009) Novel roles for A-type lamins in telomere biology and the DNA damage response pathway. *EMBO J*, **28**, 2414-2427.
173. Hozak, P., Sasseville, A.M., Raymond, Y. and Cook, P.R. (1995) Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells. *Journal of Cell Science*, **108**, 635-644.
174. Ohki, R., Tsurimoto, T. and Ishikawa, F. (2001) In Vitro Reconstitution of the End Replication Problem. *Molecular and Cellular Biology*, **21**, 5753-5766.
175. Huffman, K.E., Levene, S.D., Tesmer, V.M., Shay, J.W. and Wright, W.E. (2000) Telomere Shortening Is Proportional to the Size of the G-rich Telomeric 3′-Overhang. *Journal of Biological Chemistry*, **275**, 19719-19722.
176. Wu, P., van Overbeek, M., Rooney, S. and de Lange, T. (2010) Apollo Contributes to G Overhang Maintenance and Protects Leading-End Telomeres. *Molecular Cell*, **39**, 606-617.

177. Lam, Y.C., Akhter, S., Gu, P., Ye, J., Poulet, A., Giraud-Panis, M.J., Bailey, S.M., Gilson, E., Legerski, R.J. and Chang, S. (2010) SNMIB/Apollo protects leading-strand telomeres against NHEJ-mediated repair. **29**, 2230-2241.
178. van Overbeek, M. and de Lange, T. (2006) Apollo, an Artemis-related nuclease, interacts with TRF2 and protects human telomeres in S phase. *Curr Biol*, **16**, 1295-1302.
179. Lenain, C., Bauwens, S., Amiard, S., Brunori, M., Giraud-Panis, M.J. and Gilson, E. (2006) The Apollo 5' exonuclease functions together with TRF2 to protect telomeres from DNA repair. *Curr Biol*, **16**, 1303-1310.
180. Wu, P., Takai, H. and de Lange, T. (2012) Telomeric 3′ Overhangs Derive from Resection by Exo1 and Apollo and Fill-In by POT1b-Associated CST. *Cell*, **150**, 39-52.
181. Sfeir, A.J., Chai, W., Shay, J.W. and Wright, W.E. (2005) Telomere-end processing the terminal nucleotides of human chromosomes. *Mol Cell*, **18**, 131-138.
182. Miyake, Y., Nakamura, M., Nabetani, A., Shimamura, S., Tamura, M., Yonehara, S., Saito, M. and Ishikawa, F. (2009) RPA-like Mammalian Ctc1-Stn1-Ten1 Complex Binds to Single-Stranded DNA and Protects Telomeres Independently of the Pot1 Pathway. *Molecular Cell*, **36**, 193-206.
183. Surovtseva, Y.V., Churikov, D., Boltz, K.A., Song, X., Lamb, J.C., Warrington, R., Leehy, K., Heacock, M., Price, C.M. and Shippen, D.E. (2009) Conserved Telomere Maintenance Component 1 Interacts with STN1 and Maintains Chromosome Ends in Higher Eukaryotes. *Molecular Cell*, **36**, 207-218.
184. Dai, X., Huang, C., Bhusari, A., Sampathi, S., Schubert, K. and Chai, W. (2010) Molecular steps of G-overhang generation at human telomeres and its function in chromosome end protection. *EMBO J*, **29**, 2788-2801.
185. Hayashi, N. and Murakami, S. (2002) STM1, a gene which encodes a guanine quadruplex binding protein, interacts with CDC13 in Saccharomyces cerevisiae. *Mol Gen Genomics*, **267**, 806-813.
186. Smith, J.S., Chen, Q., Yatsunyk, L.A., Nicoludis, J.M., Garcia, M.S., Kranaster, R., Balasubramanian, S., Monchaud, D., Teulade-Fichou, M.-P., Abramowitz, L. *et al.* (2011) Rudimentary G-quadruplex–based telomere capping in Saccharomyces cerevisiae. *Nat Struct Mol Biol*, **18**, 478-485.
187. Martínez, P., Thanasoula, M., Muñoz, P., Liao, C., Tejera, A., McNees, C., Flores, J.M., Fernández-Capetillo, O., Tarsounas, M. and Blasco, M.A. (2009) Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. *Genes & Development*, **23**, 2060-2075.

188. Dvir, A., Conaway, J.W. and Conaway, R.C. (2001) Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Current Opinion in Genetics & Development*, **11**, 209-214.
189. Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. and Kornberg, R.D. (2001) Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. *Science*, **292**, 1876-1882.
190. McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M. and Bentley, D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, **385**, 357-361.
191. Nag, A., Narsinh, K. and Martinson, H.G. (2007) The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nat Struct Mol Biol*, **14**, 662-669.
192. Orphanides, G. and Reinberg, D. (2002) A Unified Theory of Gene Expression. *Cell*, **108**, 439-451.
193. Vizlin-Hodzic, D., Johansson, H., Ryme, J., Simonsson, T. and Simonsson, S. (2011) SAF-A has a role in transcriptional regulation of Oct4 in ES cells through promoter binding. *Cell Reprogram*, **13**, 13-27.
194. Kim, M.K. and Nikodem, V.M. (1999) hnRNP U Inhibits Carboxy-Terminal Domain Phosphorylation by TFIIH and Represses RNA Polymerase II Elongation. *Molecular and Cellular Biology*, **19**, 6833-6844.
195. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.
196. Robinson, P.J.J., Fairall, L., Huynh, V.A.T. and Rhodes, D. (2006) EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. *Proceedings of the National Academy of Sciences*, **103**, 6506-6511.
197. Schalch, T., Duda, S., Sargent, D.F. and Richmond, T.J. (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, **436**, 138-141.
198. Luger, K., Dechassa, M.L. and Tremethick, D.J. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol*, **13**, 436-447.
199. Kouzarides, T. (2007) Chromatin Modifications and Their Function. *Cell*, **128**, 693-705.
200. Lu, X., Simon, M.D., Chodaparambil, J.V., Hansen, J.C., Shokat, K.M. and Luger, K. (2008) The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat Struct Mol Biol*, **15**, 1122-1124.
201. Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R. and Peterson, C.L. (2006) Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science*, **311**, 844-847.

202. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. and Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, **410**, 116-120.
203. Clapier, C.R. and Cairns, B.R. (2009) The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, **78**, 273-304.
204. Eisen, J.A., Sweder, K.S. and Hanawalt, P.C. (1995) Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Research*, **23**, 2715-2723.
205. Mohrmann, L. and Verrijzer, C.P. (2005) Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, **1681**, 59-73.
206. Agalioti, T., Chen, G. and Thanos, D. (2002) Deciphering the Transcriptional Histone Acetylation Code for a Human Gene. *Cell*, **111**, 381-392.
207. Shen, W., Xu, C., Huang, W., Zhang, J., Carlson, J.E., Tu, X., Wu, J. and Shi, Y. (2007) Solution Structure of Human Brg1 Bromodomain and Its Specific Binding to Acetylated Histone Tails. *Biochemistry*, **46**, 2100-2110.
208. Lomvardas, S. and Thanos, D. (2001) Nucleosome Sliding via TBP DNA Binding In Vivo. *Cell*, **106**, 685-696.
209. Kwon, H., Imbalzano, A.N., Khavari, P.A., Kingston, R.E. and Green, M.R. (1994) Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature*, **370**, 477-481.
210. Kadam, S. and Emerson, B.M. (2003) Transcriptional Specificity of Human SWI/SNF BRG1 and BRM Chromatin Remodeling Complexes. *Molecular Cell*, **11**, 377-389.
211. Neish, A.S., Anderson, S.F., Schlegel, B.P., Wei, W. and Parvin, J.D. (1998) Factors associated with the mammalian RNA polymerase II holoenzyme. *Nucleic Acids Research*, **26**, 847-853.
212. Subtil-Rodriguez, A. and Reyes, J.C. (2010) BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo. *EMBO Rep*, **11**, 751-757.
213. Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J. and Crabtree, G.R. (2009) An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences*, **106**, 5181-5186.
214. Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K. and Crabtree, G.R. (2009) An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proceedings of the National Academy of Sciences*, **106**, 5187-5191.

215. Ho, L., Miller, E.L., Ronan, J.L., Ho, W.Q., Jothi, R. and Crabtree, G.R. (2011) esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat Cell Biol*, **13**, 903-913.
216. Takebayashi, S.-i., Lei, I., Ryba, T., Sasaki, T., Dileep, V., Battaglia, D., Gao, X., Fang, P., Fan, Y., Esteban, M. *et al.* (2013) Murine esBAF chromatin remodeling complex subunits BAF250a and Brg1 are necessary to maintain and reprogram pluripotency-specific replication timing of select replication domains. *Epigenetics & Chromatin*, **6**, 42.
217. Verma, A., Halder, K., Halder, R., Yadav, V.K., Rawal, P., Thakur, R.K., Mohd, F., Sharma, A. and Chowdhury, S. (2008) Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved cis-Regulatory Elements in Human and Related Species. *Journal of Medicinal Chemistry*, **51**, 5641-5649.
218. Zhang, C., Liu, H.-h., Zheng, K.-w., Hao, Y.-h. and Tan, Z. (2013) DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix. *Nucleic Acids Research*, **41**, 7144-7152.
219. González, V., Guo, K., Hurley, L. and Sun, D. (2009) Identification and Characterization of Nucleolin as a c-myc G-quadruplex-binding Protein. *Journal of Biological Chemistry*, **284**, 23622-23635.
220. Berberich, S.J. and Postel, E.H. (1995) PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element. *Oncogene*, **10**, 2343-2347.
221. Thakur, R.K., Kumar, P., Halder, K., Verma, A., Kar, A., Parent, J.-L., Basundra, R., Kumar, A. and Chowdhury, S. (2009) Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Research*, **37**, 172-183.
222. Sun, D. and Hurley, L.H. (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J Med Chem*, **52**, 2863-2874.
223. Raiber, E.-A., Kranaster, R., Lam, E., Nikan, M. and Balasubramanian, S. (2012) A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Research*, **40**, 1499-1508.
224. Dynan, W.S. and Tjian, R. (1983) The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell*, **35**, 79-87.
225. Chiarella, S., De Cola, A., Scaglione, G.L., Carletti, E., Graziano, V., Barcaroli, D., Lo Sterzo, C., Di Matteo, A., Di Ilio, C., Falini, B. *et al.* (2013) Nucleophosmin mutations alter its nucleolar localization

by impairing G-quadruplex binding at ribosomal DNA. *Nucleic Acids Research*, **41**, 3228-3239.

226. Belenguer, P., Baldin, V., Mathieu, C., Prats, H., Bensaid, M., Bouche, G. and Amalric, F. (1989) Protein kinase NII and the regulation of rDNA transcription in mammalian cells. *Nucleic Acids Research*, **17**, 6625-6636.
227. Peter, M., Nakagawa, J., Dorée, M., Labbé, J.C. and Nigg, E.A. (1990) Identification of major nucleolar proteins as candidate mitotic substrates of cdc2 kinase. *Cell*, **60**, 791-801.
228. Borer, R.A., Lehner, C.F., Eppenberger, H.M. and Nigg, E.A. (1989) Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell*, **56**, 379-390.
229. Okuda, M., Horn, H.F., Tarapore, P., Tokuyama, Y., Smulian, A.G., Chan, P.-K., Knudsen, E.S., Hofmann, I.A., Snyder, J.D., Bove, K.E. *et al.* (2000) Nucleophosmin/B23 Is a Target of CDK2/Cyclin E in Centrosome Duplication. *Cell*, **103**, 127-140.
230. Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C.B., Proffitt, C., Trent, K., Whitten, J.P. *et al.* (2009) Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Research*, **69**, 7653-7661.
231. Gibbons, R.J., Picketts, D.J., Villard, L. and Higgs, D.R. (1995) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with α-thalassemia (ATR-X syndrome). *Cell*, **80**, 837-845.
232. Soderberg, O., Gullberg, M., Jarvius, M., Ridderstrale, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G. *et al.* (2006) Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods*, **3**, 995-1000.
233. Förster, T. (1965) Delocalized excitation and excitation transfer. *Modern quantum chemistry*, **3**, 99-137.
234. Han, R., Li, Z., Fan, Y. and Jiang, Y. (2013) Recent Advances in Super-Resolution Fluorescence Imaging and Its Applications in Biology. *Journal of Genetics and Genomics*, **40**, 583-595.
235. Yao, Y., Wang, Q., Hao, Y.-h. and Tan, Z. (2007) An exonuclease I hydrolysis assay for evaluating G-quadruplex stabilization by small molecules. *Nucleic Acids Research*, **35**, e68.
236. Han, H., Hurley, L.H. and Salazar, M. (1999) A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Research*, **27**, 537-542.
237. Takahama, K., Kino, K., Arai, S., Kurokawa, R. and Oyoshi, T. (2011) Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS Journal*, **278**, 988-998.
238. Qureshi, M.H., Ray, S., Sewell, A.L., Basu, S. and Balci, H. (2012) Replication Protein A Unfolds G-Quadruplex Structures with

Varying Degrees of Efficiency. *The Journal of Physical Chemistry B*, **116**, 5588-5594.

239. Montgomery, M.K., Xu, S. and Fire, A. (1998) RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, **95**, 15502-15507.
240. Hammond, S.M., Bernstein, E., Beach, D. and Hannon, G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, **404**, 293-296.
241. Romig, H., Fackelmayer, F.O., Renz, A., Ramsperger, U. and Richter, A. (1992) Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *EMBO J*, **11**, 3431-3440.
242. Reyes, J.C., Muchardt, C. and Yaniv, M. (1997) Components of the Human SWI/SNF Complex Are Enriched in Active Chromatin and Are Associated with the Nuclear Matrix. *The Journal of Cell Biology*, **137**, 263-274.
243. Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Randazzo, F., Metzger, D., Chambon, P., Crabtree, G. *et al.* (2000) A Brg1 Null Mutation in the Mouse Reveals Functional Differences among Mammalian SWI/SNF Complexes. *Molecular Cell*, **6**, 1287-1295.
244. Roshon, M. and Ruley, H.E. (2005) Hypomorphic mutation in hnRNP U results in post-implantation lethality. *Transgenic Res*, **14**, 179-192.
245. Kim, S.-I., Bresnick, E.H. and Bultman, S.J. (2009) BRG1 directly regulates nucleosome structure and chromatin looping of the α globin locus to activate transcription. *Nucleic Acids Research*, **37**, 6019-6027.
246. Koziol, M.J., Garrett, N. and Gurdon, J.B. (2007) Tpt1 Activates Transcription of oct4 and nanog in Transplanted Somatic Nuclei. *Current Biology*, **17**, 801-807.
247. Tani, T., Shimada, H., Kato, Y. and Tsunoda, Y. (2007) Bovine Oocytes with the Potential to Reprogram Somatic Cell Nuclei Have a Unique 23-kDa Protein, Phosphorylated Transcriptionally Controlled Tumor Protein (TCTP). *Cloning and Stem Cells*, **9**, 267-280.
248. Baharvand, H., Fathi, A., Gourabi, H., Mollamohammadi, S. and Salekdeh, G.H. (2007) Identification of Mouse Embryonic Stem Cell-Associated Proteins. *Journal of Proteome Research*, **7**, 412-423.
249. Gachet, Y., Tournier, S., Lee, M., Lazaris-Karatzas, A., Poulton, T. and Bommer, U.A. (1999) The growth-related, translationally controlled protein P23 has properties of a tubulin binding protein and associates transiently with microtubules during the cell cycle. *Journal of Cell Science*, **112**, 1257-1271.

250. Ugrinova, I., Monier, K., Ivaldi, C., Thiry, M., Storck, S., Mongelard, F. and Bouvet, P. (2007) Inactivation of nucleolin leads to nucleolar disruption, cell cycle arrest and defects in centrosome duplication. *BMC Mol Biol*, **8**, 66.
251. Ma, N., Matsunaga, S., Takata, H., Ono-Maniwa, R., Uchiyama, S. and Fukui, K. (2007) Nucleolin functions in nucleolus formation and chromosome congression. *Journal of Cell Science*, **120**, 2091-2105.
252. Scholer, H.R., Dressler, G.R., Balling, R., Rohdewohld, H. and Gruss, P. (1990) Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *EMBO J*, **9**, 2185-2195.
253. Yang, T.H., Tsai, W.H., Lee, Y.M., Lei, H.Y., Lai, M.Y., Chen, D.S., Yeh, N.H. and Lee, S.C. (1994) Purification and characterization of nucleolin and its identification as a transcription repressor. *Mol Cell Biol*, **14**, 6068-6074.
254. Hanakahi, L.A., Sun, H. and Maizels, N. (1999) High Affinity Interactions of Nucleolin with G-G-paired rDNA. *Journal of Biological Chemistry*, **274**, 15908-15912.
255. Ho, L. and Crabtree, G.R. (2010) Chromatin remodelling during development. *Nature*, **463**, 474-484.