DEVELOPMENT OF EMBRYONIC STEM CELLS EXPRESSING ENDOGENOUS LEVELS OF A FLUORESCENT PROTEIN FUSED TO THE TELOMERE BINDING PROTEIN TRF1

by

SHELLEY BONNIE MILLER

B.Sc., The University of British Columbia, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2008

Abstract

Telomeres are the repetitive DNA sequence and associated proteins found at the ends of linear chromosomes. They have a role in biological processes including meiosis and aging as well as implications in a number of genomic instability disorders and cancers. Telomeres maintain genomic stability by protecting chromosome ends from terminal fusions and misidentification as DNA damage sites. Their wide range of functions has resulted in an increased interest in developing tools to study the dynamics of telomeres in live cells. To do this, current studies use the ubiquitously expressed protein Telomere Repeat Factor 1 (TRF1) tagged with a fluorescent protein. TRF1 is a negative regulator of telomere length that binds exclusively to telomere repeats. Over-expression of the fluorescent protein fused to TRF1 has been a useful tool to track telomere movement. The foci formed by the tagged TRF1 protein accurately represent the number of telomeres expected in the cells and the localization is maintained throughout the cell cycle. A caveat with this system is that over-expression of TRF1 leads to accelerated telomere shortening, as well as replication defects that can stall telomere replication. These caveats make it difficult to draw conclusions about telomere dynamics based solely on observations of cells over-expressing fluorescently tagged TRF1. To eliminate problems associated with protein over-expression, I have tried to develop knock-in embryonic stem (ES) cells expressing fluorescently tagged TRF1 from the endogenous *Trf1* promoter. To do this, I have used a recombineering technique using Bacterial Artificial Chromosomes (BACs). BAC recombineering allows for the direct knock-in of a fluorescent tag into the mouse *Trf1* gene locus. Genetic constructs with the correct sequence inserts have been obtained and have been used for transfection of ES cells. While no correctly targeted ES cells have been identified so far, the expectation is that ES cell lines with correctly targeted fluorescently tagged TRF1 will be obtained in the near future. Such lines will be used to study telomere dynamics in ES cells, differentiated cells generated from ES cells, as well as to generate mice.

Table of Contents

.

List of Tables

List of Figures

List of Abbreviations

TRF1: Telomere Repeat Factor 1

TRF2: Telomere Repeat Factor 2

POT1: Protection of Telomeres 1

PNA: peptide nucleic acid

GFP: green fluorescent protein

ES cell: embryonic stem cell

BACs: bacterial artificial chromosomes

dsDNA: double-strand DNA

neo^r: Neomycin resistance gene

LB: Luria-Bertani media

DOG: 2-deoxy-galactose

DMEM: Dulbecco's Modified Eagle Medium

FCS: Fetal Calf Serum

LIF: leukemia inhibitory factor

FACS: fluorescence activated cell sorting

Acknowledgments

I would like to thank my supervisor Dr. Peter Lansdorp for giving me the opportunity to work on this project in such a great environment. I would also like to thank Dr. Evert-Jan Uringa for his help and guidance with the recombineering work. In addition, I would like to express thanks to Dr. Ester Falconer, Dr. Geraldine Aubert, and Michelle Decker for all their help, suggestions, and time spent editing. Thank you to the entire Lansdorp lab for their help and encouragement throughout my time in the lab. Thank you to the Flow Cytometry staff in the Terry Fox Lab for their help with cell sorting. Finally, I would like to express gratitude to my committee members, Dr. Louis Lefebvre, Dr. Wendy Robinson, and Dr. Ann Rose for their suggestions and insight.

Dedication

I would like to thank my fiancé, Brian Weisser, for his never-ending love, support and patience. Thank you to my grandpa, Otto Miller, for his constant encouragement. I would also like to thank mom, Thelma Weisser, for her support.

Always in loving memory of my dad, Toto Miller, and my grandma, Jeanne Miller.

Chapter 1: Introduction

1.1 Telomeres

1.1.1 Telomere structure

Eukaryotic chromosomes face challenges in maintaining stability due to their linearity. Not only do the natural ends of chromosomes need to be protected from breakdown and degradation, but chromosome ends also have to avoid being recognized and processed as double strand breaks.[1] Protection from these processes is achieved by specialized nucleoprotein structures called telomeres.

Telomeric DNA is composed of tandem arrays of G-rich repeats. In vertebrates, the telomere repetitive sequence is $(TTAGGG)_n$. The number of repeats, and therefore telomere length, can vary greatly between individual chromosomes and cells.[2] While most telomeric DNA is in a double stranded conformation, the extreme end of the chromosome forms a 3' single strand overhang between 50 and 300 nucleotides long.[1, 3] Associated with this sequence is a six protein complex is called the telosome or shelterin due to its implied contribution to telomere protection (Figure 1A).[4, 5]

The first protein identified from the shelterin complex was Telomere Repeat Factor 1 (TRF1) due to its *in vitro* specificity for double-stranded TTAGGG repeats.[6] TRF1 is believed to be a negative regulator of telomere length because over-expression leads to a gradual decrease in length while deletion leads to an abnormal elongation of telomere length.[7] Shortly after, Telomere Repeat Factor 2 (TRF2) was identified through sequence homology as a paralog of TRF1.[8] TRF2 also binds telomere TTAGGG repeats and helps protect the 3' overhang from nucleases and DNA damage detection.[7] The

protein TIN2 tethers both TRF1 and TRF2 as well as binds the additional shelterin protein, TPP1. TPP1 serves to recruit another protein, Protection of Telomeres 1 (POT1), to the telomeric complex.[9] POT1 acts uniquely in this complex by binding the G-rich single-stranded overhang (G-tail) and averting DNA damage signaling at chromosome ends.[10] The last shelterin protein associated with the complex is Rap1. Rap1 is recruited to telomeres by TRF2 and is implicated in playing a role in the regulation of telomere length.[11]

Telomere length affects the number of shelterin complexes bound to the repeats. This acts as a length sensing mechanism since longer telomeres have more shelterin proteins bound compared to shorter telomeres.[12] In addition, these complexes are proposed to affect the structure of telomeric DNA by forming T-loops (Figure 1B). T-loops occur when the 3' single-strand overhang invades the duplex telomeric repeat array and forms a lariat structure.[5] TRF1 can induce bending and looping of telomeric DNA while TRF2 may facilitate the invasion of the G-tail into the duplex telomeric repeats.[7, 12] T-loop structures were first identified using electron microscopy of telomeric restriction fragments from both human and mouse cells and provide evidence supporting a T-loop mediated mechanism of protection for telomeres.[13] By sequestering chromosome ends into T-loops, telomeres are not recognized as double strand DNA breaks and thus prevents cellular activities from the DNA damage machinery or nucleases.[13]

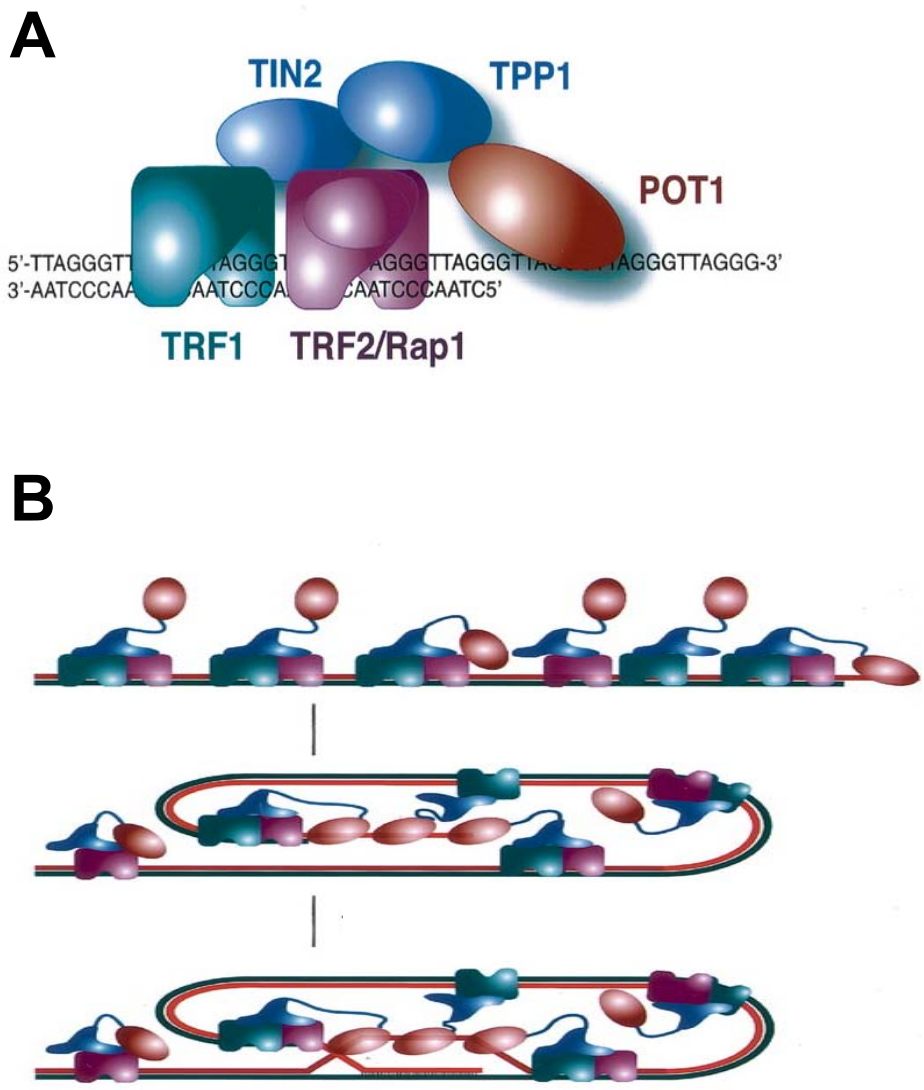

Figure 1. Representation of mammalian telomeres and associated proteins. A. Telomere DNA sequence is associated with the shelterin complex which is composed of six proteins, TRF1, TRF2, Rap1, TIN2, TPP1, and POT1. B. Shelterin complexes localize along the telomere and facilitate a T-loop formation where the telomere bends back and embeds the single strand overhang in the double stranded telomeric DNA (Reprinted by permission from CSHL Press: Genes and Development, de Lange, T., 2005).[5]

1.1.2 Telomere function

The unique structure of telomeres allows them to perform their protective function at chromosome ends. Broken chromosomes or free DNA ends are susceptible to end-to-end fusions and exonucleolytic degradation. Telomeres deter end-to-end chromosome fusions and thus prevent the formation of unstable dicentric chromosomes.[14] In addition, the telomeric protective structure prevents the loss of genetic information from sub-telomeric regions that can occur through stochastic processes such as oxidative damage and replication error.[14] Telomere loss also occurs through the "end-replication problem". During lagging strand synthesis, the terminal primer is degraded, resulting in a 5' gap at both ends of the chromosome. All ends are further processed by a 5' to 3' exonuclease which degrades 130-210 nucleotides, resulting in progressive telomere shortening with each replication.[15] Once the telomeres reach a critical length where genes are potentially compromised, cells will undergo replicative senescence or apoptosis. If the cell bypasses this checkpoint, chromosomal instability or end-to-end fusions can occur.[15] The enzyme telomerase acts by adding repetitive telomere sequence to chromosome ends to maintain or elongate telomeres.[15] With the exception of B cells and stem cells of the germline all somatic cells including hematopoietic stem cells and T lymphocytes appear to loose telomere repeats with each cell division despite expressing limiting levels of telomerase.[1]

In addition to protecting sub-telomeric genetic material from degradation, telomeres are also important in distinguishing chromosome ends from DNA double strand breaks and therefore preventing DNA damage signaling.[5] Telomere erosion or mutations in shelterin proteins can result in unprotected telomeres recruiting DNA

damage response factors leading to activation of ATM kinase and a p53-dependent G1/S arrest inducing either apoptosis or senescence.[16]

1.1.3 Telomeres and meiosis

During prophase I of meiosis, telomeres have a specific pattern of movement in preparation for cell division. During the leptotene-zygotene stages, telomeres migrate along the nuclear envelope to the centrosome. Here, they become tightly clustered and form a "bouquet" structure with the telomeres clustered at the centrosome and the chromosomes looping out.[17] This process facilitates the alignment, pairing, recombination and segregation of homologous chromosomes.[18] Telomere dysfunction resulting from deleting shelterin proteins can disrupt these processes leading to serious biological consequences including reduced homologous recombination and increased mis-segregation. This can lead to sterility, aneuploidy, failed implantation and miscarriage.[17, 19] The number and severity of biological problems associated with telomere dysfunction during meiosis illustrates the importance of fully examining telomere function throughout the process of gametogenesis.

1.1.4 Telomeres and aging

Telomeres shorten progressively with each cell division as well as stochastically as detailed in section 1.1.2. Progressive telomere loss is an important timing mechanism linked to replicative senescence, which is characteristic of "aging" both in cell culture and *in vivo*.[20] The telomere hypothesis of aging is based on a number of observations in human cells. These include the observations that telomeres are shorter in somatic tissues

of older individuals compared to younger individuals, as well as being shorter in somatic cells than in germline cells.[20] Telomeres of cells taken from young individuals progressively shorten when grown in cell culture. By contrast, experimentally elongating telomeres through telomerase expression extends the proliferative capacity of cultured human cells.[20] In addition, children born with progeria (accelerated aging syndromes) have shorter telomeres in some somatic tissues than age-matched controls.[21] While further examination is needed to determine if telomere loss leading to cellular senescence is actually representative of biological aging, mounting evidence suggests that genetically impaired telomere maintenance contributes to both accelerated and normal aging.[22]

1.1.5 Telomeres and cancer

Telomere dysfunction can result in end fusion events resulting in dicentric chromosomes that can lead to aberrations and genome instability. Occasionally cells can survive this genomic destabilization and accumulate genetic alterations that lead to neoplasia.[22] In addition, bypassing replicative senescence is thought to be a critical rate-limiting step in the progression of malignancies. Most malignant tumors bypass senescence to acquire the unlimited proliferative capacity required for established tumors.[23] Cancer cells achieve this by restoring telomere length, primarily by activating telomerase. Telomerase provides the critical functions of suppressing chromosomal instability and allowing for unlimited replication or "immortalization" necessary in cancer cells.[24] In fact, 80%-90% of tumors show activation of telomerase expression making it an important drug target for cancer. It is a relatively specific cancer target as

somatic cells express little or no telomerase for most of their lifespan and generally have longer telomeres than those in tumour cells.[25]

1.1.6 Visualizing telomere dynamics

The role that telomeres play in normal biological processes, diseases and genomic instability makes them an important focus of study. Therefore it has become of increasing interest to develop tools to visually examine telomere dynamics. For example, using telomere specific peptide nucleic acid (PNA) probes to identify telomeres on fixed cells has allowed examination into their specialized role in meiosis and proper chromosome segregation.[26] In addition, PNA probes used with live-cell imaging has shown that telomeres move dynamically within the cell and associate both with other telomeres and additional proteins.[27] Previous studies have used various TRF1 fusion proteins to visually examine telomere dynamics due to TRF1 specificity for telomere repeats and ubiquitous expression.[28] A FLAG-tagged TRF1 expression vector with antibodies directed against the FLAG peptide showed that murine TRF1 localizes specifically to telomeres and this localization is maintained and can be followed throughout the cell cycle.[28] In addition, anti-TRF1 antibodies are used to distinguish telomeres when determining if proteins co-localize to telomeres.[29, 30] A final method of using TRF1 to visualize telomeres is by over-expressing a green fluorescent protein (GFP)-TRF1 fusion protein. This method has been used to show the dynamic nature of TRF1 binding as well as co-localization with other fluorescent-tagged telomere associating proteins.[31] In fact, GFP-TRF1 experiments have shown that telomere localization is recovered within a minute of photobleaching therefore demonstrating the reversible binding of TRF1 to telomeres.[31] The specificity of

TRF1 to telomeric DNA makes it a good candidate for visualizing telomeres while studying their dynamic behavior.

1.2 Telomere Repeat Factor 1

1.2.1 TRF1 structure

TRF1 is expressed in all cells of humans, monkeys, rodents, and other vertebrates.[32] Its expression is essential for survival, as a targeted deletion of exon 1 in mice results in embryonic lethality.[33] The molecular structure of TRF1 includes a 50 amino acid C-terminal Myb DNA-binding domain that specifically binds double-stranded TTAGGG repeats (Figure 2A).[6, 34] It also contains a ~200 amino acid TRF-specific domain, called the dimerization domain, that mediates homodimerization which is required for binding to telomeric DNA.[35] The acidic N-terminus domain of TRF1 binds to the regulatory proteins tankyrase 1 and 2 which can modify TRF1 to impede its DNA binding activity or remove TRF1 from telomeres and promote its degradation.[9] Although the spatial arrangement of the Myb DNA-binding domains within the homodimer remains to be determined, they act cooperatively for functional DNA binding. Possible binding scenarios include the monomer DNA binding domains localizing side by side or some distance from each other, either in the same or opposite orientations (Figure 2B).[35] While the dimerization domain and the Myb DNA binding domain have maintained high sequence identity between humans and mice, there is considerable divergence in other regions of the protein suggesting that telomeric proteins evolve relatively quickly.[28]

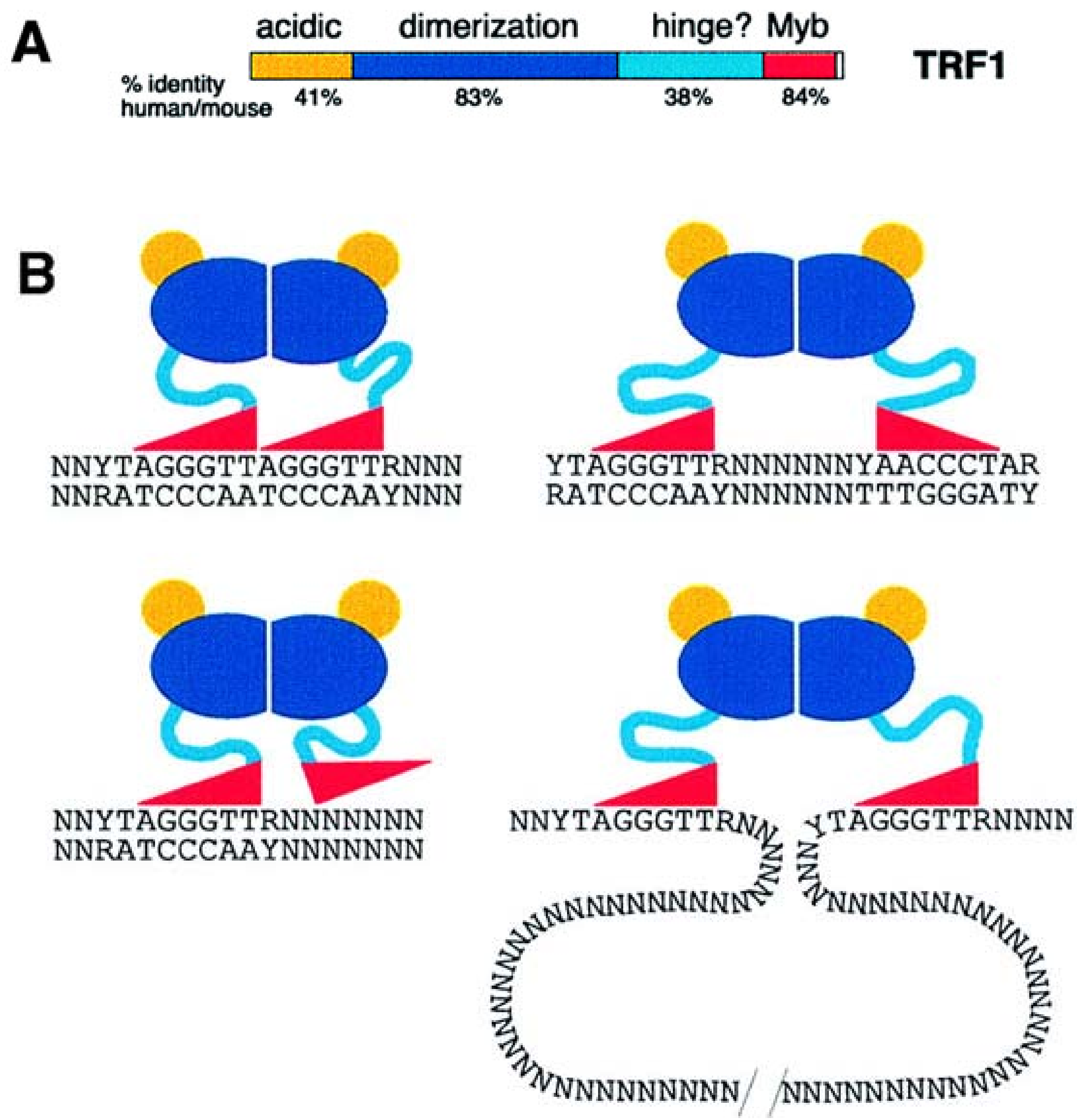

Figure 2. Representation of TRF1 protein structure and proposed models of telomere binding. A. TRF1 consists of three domains, the acidic domain (yellow) recognized by the protein tankyrase which regulates telomere association, the TRF dimerization domain (blue) for homodimer association, and the Myb DNA binding domain (red). B. Potential binding scenarios of the TRF1 homodimer. The DNA binding domains can associate with TTAGGG repeats in the same or opposite orientation either side by side or some distance from each other (Reprinted by permission from Macmillan Publishers Ltd: The EMBO Journal, Bianchi et al, copyright 1999).[35]

1.2.2 TRF1 function

Telomere length homeostasis is maintained by highly dynamic binding of TRF1 and TRF2.[31] The short association time of TRF1 (half time of ~8 seconds as determined by fluorescence recovery after photobleaching analysis) can allow quick adjustments to telomere length by regulating access for telomerase-mediated elongation or nuclease-mediated shortening of telomeric DNA.[31] In addition, this dynamic association is necessary for successful DNA replication as bound TRF1 stalls replication forks and over-expression of TRF1 leads to telomere replication defects and delayed S phase exit.[36] Over-expression of TRF1 induces telomere shortening even in the presence of telomerase, while the expression of a dominant-negative TRF1 results in elongation of telomeres.[37] This supports the role of TRF1 as a negative regulator of telomere length by a *cis*-acting mechanism. By direct binding of TRF1 to telomere repeats, cells can monitor and regulate the length of individual telomeres so that longer telomeres recruit more TRF1 thereby exerting stronger negative feedback on telomerase and inhibiting the enzyme.[30]

TRF1 has the ability to induce bending, looping, and pairing of duplex telomeric DNA *in vitro*.[32] These properties could facilitate the folding back of the telomere in preparation of T-loop formation with the assistance of other telomeric proteins such as TIN2, TPP1, and POT1.[4, 12] Iwano et al (2004) demonstrated that murine embryonic stem cells deficient in TRF1 expression resulted in a decrease of TRF2 binding as well as a loss of TIN2 association suggesting that TRF1 is essential for proper shelterin complex formation.[38]

1.3 Project purpose

TRF1 can be utilized as an important tool to visualize and study telomeres. The fusion of a fluorescent protein to TRF1 in an over-expression vector can be used to track telomeres using live-cell imaging. This method accurately labels telomeres and localization is maintained throughout the cell cycle (Figure 3, personal communication with Kathleen Lisaingo). Because telomeres move dynamically within a cell it is useful to track their movement and associations to elucidate their specialized functions. In addition, accumulation of TRF1 at chromosome ends correlates with telomere length where long telomeres would recruit more TRF1, blocking access to telomerase.[12, 29] Therefore, levels of fluorescence also correlate to telomere length allowing studies of telomere length dynamics.[39]

A caveat with this system is that over-expression of TRF1 leads to accelerated telomere shortening, as well as replication defects that can stall S phase during cell division.[36, 37] Therefore, the artificial skewing of telomere protein equilibrium makes it difficult to accurately study telomere dynamics in cells that over-express TRF1. To eliminate problems associated with protein over-expression, I am developing knock-in embryonic stem (ES) cells expressing endogenous levels of fluorescently-tagged TRF1. To develop these cell lines, I am using the relatively novel technique of recombineering to knock-in fluorescently tagged *Trf1* at the mouse *Trf1* gene locus, thereby maintaining endogenous regulation of tagged TRF1 expression.

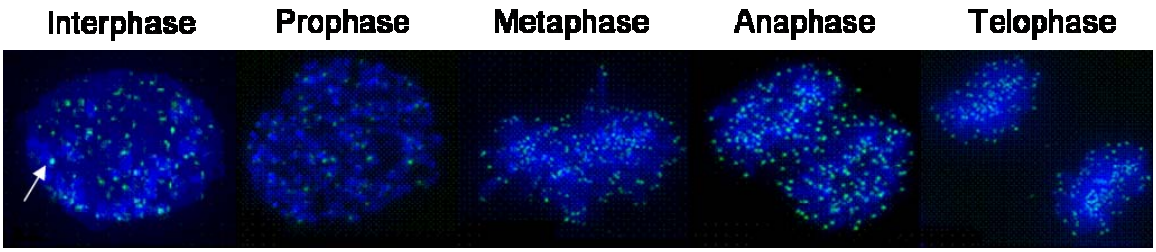

Figure 3. Venus-TRF1 accurately labels telomeres and this localization is maintained throughout the cell cycle. Murine ES cells are transfected with an over-expression vector containing TRF1 fused to a Venus fluorescent protein. DNA is visualized by DAPI staining in blue while Venus-TRF1 binding to telomeres forms green foci (white arrow). The number of foci accurately depicts the number of telomeres within the cell and this localization is maintained throughout mitosis. (Figure provided by Kathleen Lisaingo)

1.4 Recombineering

A novel form of genetic engineering, called recombinogenic engineering or recombineering, has recently been developed allowing modification or subcloning of genomic DNA in bacterial artificial chromosomes (BACs) and subsequent targeting into the host cell genome.[40] This method has a number of advantages over traditional genetic engineering. First, BACs can be altered via recombination without the need for restriction enzymes or DNA ligases which eliminates the need to have or to create appropriate and unique cut sites in constructs. Secondly, recombineering allows the manipulation of large segments of DNA which is difficult using standard recombinant DNA techniques. Lastly, this technique is efficient and greatly decreases the time it takes to create transgenic mouse models.[40]

This efficient homologous recombination system is made possible by the use of strains of *E. coli* expressing phage-encoded proteins, such as the *Red* genes of bacteriophage λ. These include a 5'-3' exonuclease that produces 3' overhangs in linear double-stranded DNA (dsDNA) fragments, a pairing protein that binds to the 3' overhang promoting annealing to its complimentary DNA strand on the BAC, and *gam* which inhibits RecBCD exonuclease activity and prevents degradation of linear dsDNA.[41, 42]

Generally recombineering includes amplifying a desired DNA segment (such as a fluorescent tag) by PCR with primers containing short arms of homology to the BAC targeting region (Figure 4A). The purified PCR product is transformed into bacterial cells containing the BAC of interest and will generate a recombinant *in vivo* thereby inserting the PCR fragment into the target region. Recombinants can be detected by selection, counter selection, or by direct screening. The modified region of the BAC is transferred to a retrieval vector which contains homologous arms flanking the BAC modified region, via recombination (Figure 4B). This retrieval vector can then be introduced into ES cells or other cells of interest for genomic targeting.[40] This creates a knock-in cell line expressing endogenous levels of the tagged gene of interest.

Due to its efficiency and advantages, recombineering is becoming a useful alternative to traditional genetic engineering. Combined with other procedures such as generating conditional knockout mutations, recombineering offers new opportunities for elucidating gene function in a whole animal, as well as for creating mouse models of human disease and for testing gene therapy protocols.[41, 42]

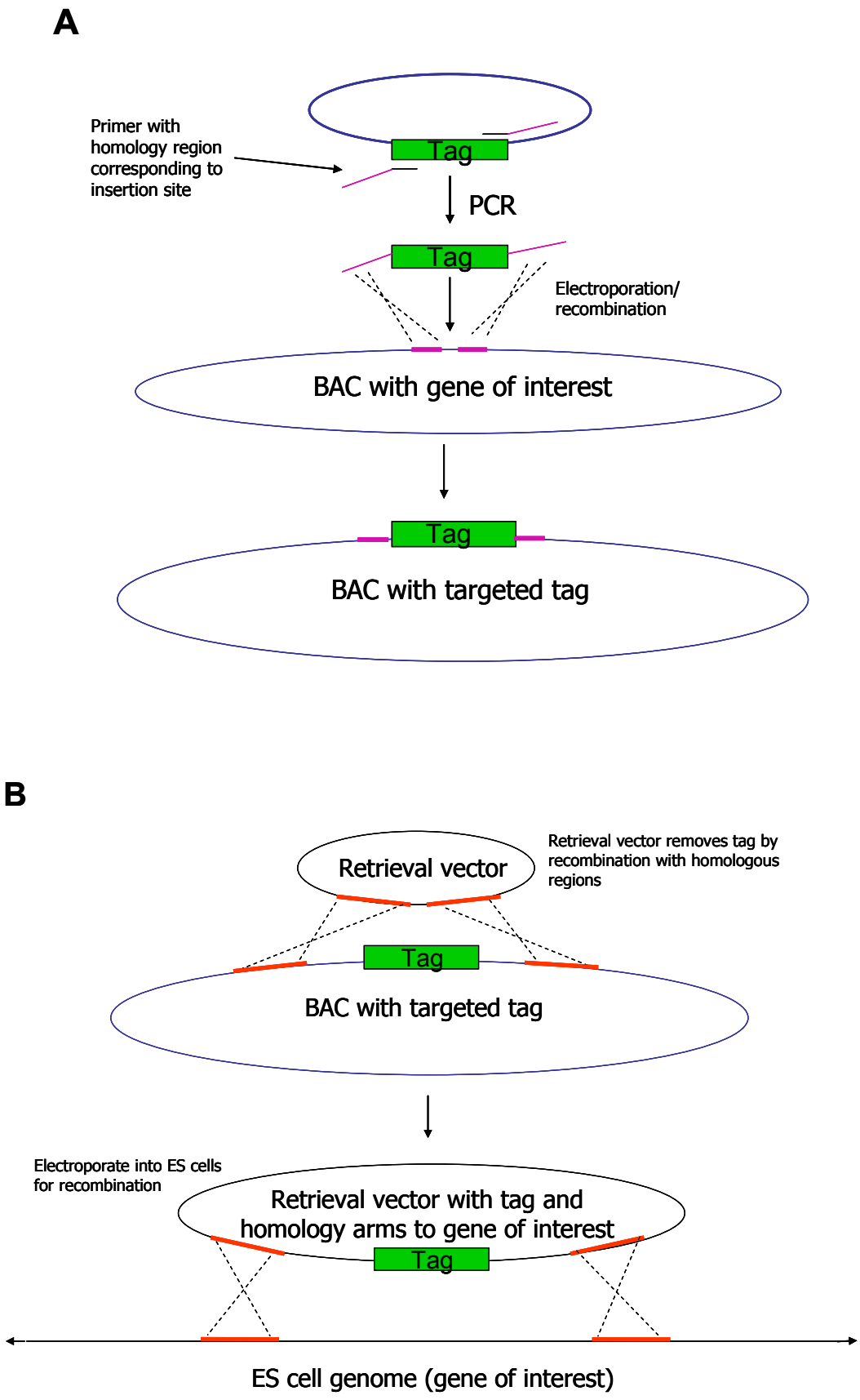
A
Primer with homology region corresponding to insertion site
Tag
PCR
Tag
Electroporation/ recombination
BAC with gene of interest
Tag
BAC with targeted tag
B
Retrieval vector
Retrieval vector removes tag by recombination with homologous regions
Tag
BAC with targeted tag
Electroporate into ES cells for recombination
Retrieval vector with tag and homology arms to gene of interest
Tag
ES cell genome (gene of interest)

Figure 4. Recombineering strategy for targeting constructs into the ES cell genome. A. The first step in recombineering is to PCR-amplify the desired Tag from a vector with primers containing short arms of homology to the gene of interest. This PCR product is then electroporated into bacteria containing BACs that have the gene of interest. Recombination results in the insertion of the Tag into the target site within the gene of interest, in a specific location. B. To target the Tag to the ES cell genome, a retrieval vector is used. The retrieval vector is constructed so it contains longer regions of homology corresponding to gene sequence upstream and downstream of the Tag insertion. The retrieval vector is electroporated into bacteria containing the BAC with the tagged gene of interest. Recombination occurs between the BAC and the retrieval vector within the homologous regions resulting in the incorporation of the Tag and target gene homologous arms into the retrieval vector. The retrieval vector with the Tag is then electroporated into ES cells where recombination will insert the Tag flanked by homology regions into the target site of the gene of interest within the ES cell genome.

Chapter 2: Materials and Methods

2.1 Development of TRF1 knock-in cell lines

2.1.1 Trf1 knock-in strategy for BAC recombineering

Two TRF1 knock-in ES cell lines are being developed. One cell line will have a Tag containing a Venus fluorescence marker and FLAG Tag targeted to the 5' region of TRF1 (called the 5' construct) while the other will have the same Tag targeted to the 3' region of the gene (called the 3' construct). Both of these constructs are made with the following recombineering steps in the *E. coli* bacterial strain SW102 with a chloramphenicol resistant BAC containing the *Mus musculus* gene *Trf1*. The first step is to insert a Neomycin resistance gene (neo^r) into the 3' or 5' construct. This gene has both a prokaryotic promoter for kanamycin selection in bacteria, as well as a eukaryotic promoter for geneticin selection in ES cells. To insert this gene, neo^r was PCR amplified with primers containing homologous arms to the *Trf1* insert site (see section 2.2, program 1 for details). For the 5' construct the primers 5'NeoF and 5'NeoR were used while the primers 3'NeoF and 3'NeoR were used for the 3' construct (see Table 1 for all primer sequences). The PCR product was then electroporated for the first recombineering step (Figure 4A) as described in section 2.3. Following recombineering, 100µl of the bacteria was plated on LB agar with 10µg/ml kanamycin and incubated at 32˚C for 2 days. Twelve colonies of each of the 5' and 3' constructs from this first recombineering step were picked and BACs were isolated according to section 2.4 and used for PCR analysis. PCR was used to verify proper neo^r targeting by amplifying across the insert region. For the 5' construct, primers Int1NeoF and Int1NeoR were used while Int9NeoF and

Int9NeoR were used for the 3' construct. PCR was performed as described in section 2.2 with program 2. Bacteria from colonies positive for insertion were used in the following recombineering step.

The second step in developing these constructs is to insert the *galK* gene into the future target site of the Venus Tag. This provides a selection step for future recombination events where BACs will recombine and swap the Venus Tag in place of *galK* allowing colonies to grow on plates containing 2-deoxy-galactose (DOG). Bacteria retaining BACs with *galK* will not survive on DOG plates.[41] The *galK* inserts were PCR amplified with primers containing homology arms to the desired insertion site in the *Trf1* gene. PCR reactions were performed as described in section 2.2 (program 1) with the primers 5'GalKF and 5'GalKR (5' construct) and 3'GalKF and 3'GalKR (3' construct). PCR products were then electroporated for the second recombineering step as described in section 2.3. Following recombineering, 100µl of bacteria was plated in 1:1, 1:10 and 1:100 serial dilutions on M63 minimal media plates (1L: 200ml 5X M63 media (2g $(NH_4)_2$ SO_4, 13.6g KH_2PO_4, 0.5mg $FeSO_4$-$7H_2O$,pH7), 1ml 1M $MgSO_4$-$7H_2O$, 0.2% galactose, 1mg biotin, 45mg L-leucine, 12.5µg/ml chloramphenicol, 10g agar). Following incubation at 32˚C for 4 days, 4 colonies for each of the 5' and 3' constructs were streaked onto MacConkey plates (MacConkey agar (Sigma), 0.2% galactose, 12.5µg/ml chloramphenicol). Six single bright red colonies (GalK+) were chosen for each of the 5' and 3' constructs and BACs were isolated according to section 2.4 and used for PCR analysis. PCR was used to verify proper *galK* targeting by amplifying across the insert region. For the 5' construct primers Ex1TagF and Ex1TagR were used while Ex10TagF and Ex10TagR were used for the 3' construct. PCR was performed as described in

section 2.2 with program 2. Bacteria from colonies positive for insertion were used in the following recombineering step.

The final step in developing the 5' and 3' constructs is to recombine these *galK* and neo^r containing BACs with the Venus Tags, thereby exchanging the *galK* gene with the Venus Tag. Tags for the 5' and 3' constructs were PCR amplified with primers containing homology arms to the desired insertion site in the *Trf1* gene. PCR reactions were performed as described in section 2.2 (program 1) with the primers 5'TagF and 5'TagR (5' construct) and 3'TagF and 3'TagR (3' construct). PCR products were then electroporated for recombineering as described in section 2.3 except in this case, recovery of bacteria was in 10ml LB in a 32˚C shaking waterbath for 4.5hrs. Following recombineering, 100µl of bacteria was plated in 1:1, 1:10 and 1:100 serial dilutions on M63 minimal media plates (as described previously with 0.2% glycerol instead of galactose and 0.2% DOG). Following incubation at 32˚C for 4 days, 12 colonies for each of the 5' and 3' constructs were chosen and BACs were isolated according to section 2.4 and used for sequence analysis. PCR was performed across the insert region with the 5' construct primers Ex1TagF and Ex1TagR while Ex10TagF and Ex10TagR were used for the 3' construct. PCR was performed as described in section 2.2 with program 2 and purified PCR products were sequenced to verify Tag integrity and conservation of open reading frame (Nucleic Acid Protein Service Unit, Vancouver). One colony for each of the 5' and 3' constructs where the BACs showed the correct Tag sequence was used for targeting to the ES cell genome.

Table 1. PCR primers used in *Trf1* knock-in ES cell development.

Name	Sequence
5'NeoF	5'-GTCTTCTGGCCTGTGTGTTGGTGAGAGACATCCAGGAATATGGTTGGCTT***CCTAGG***CGAATTCCTGCAGCCCAATTCCGATC-3'
5'NeoR	5'-GTGTTTGAACAAACACAAAGATACACTTGGCAGTGAATATGTACACTGGTG CTCTAGAACTAGTGGATCCCCTCGAGG-3'
3'NeoF	5'-CTTACTCCATTGAGGGAAAAAGTAAATGCCTTTCAGAAGTAAACCCTTCACATAC***GTTAAC***CGAATTCCTGCAGCCCAATTCCGATC-3'
3'NeoR	5'-GTAATCTTTGTGTGTGTGAAGGAGAAAAGCTTGTTCACATTTGTAATCTTTGTGCTCTAGAACTAGTGGATCCCCTCGAGG-3'
5'GalKF	5'-CAGCGCACGGCGCCAGCTGAGGCACGGCGAGCGCTTTCGGTTTAACCCTG TTGACAATTAATCATCGGCATAG-3'
5'GalKR	5'-CCAGCCCTCACGGCTCGGCGCGTCCCGGGCCGCTGAGGAGACCGTCTCCGCTCAGCACTGTCCTGCTCC-3'
3'GalKF	5'-GTCATGTTAAAAGATAGATGGAGAACAATGAAGAGACTGAAACTGATTAGCCCTGTTGACAATTAATCATCGGCATAG-3'
3'GalKR	5'-GTATTTATCCTTGTATTAAGCTGAATCAAACCCATCCAGCCTCCAGTGTCTCATCAGCACTGTCCTGCTCC-3'
5'TagF	5'-CAGCGCACGGCGCCAGCTGAGGCACGGCGAGCGCTTTCGGTTTAACATGGACTACAAAGACCATGACGGTGATTATAAAGATCATG-3'
5'TagR	5'-CCAGCCCTCACGGCTCGGCGCGTCCCGGGCCGCTGAGGAGACCGTCTCCGCGAATTCGCCAGAACCAGCAGCG-3'
3'TagF	5'-GTCATGTTAAAAGATAGATGGAGAACAATGAAGAGACTGAAACTGATTAGCGGATCCGCTGGCTCCGCTGC-3'
3'TagR	5'-GTATTTATCCTTGTATTAAGCTGAATCAAACCCATCCAGCCTCCAGTGTCTCATTACTTGTCATCGTCATCCTTGTAGTCGATG-3'
Int1NeoF	5'-CACAGACAGACGTGCAGG-3'
Int1NeoR	5'-GGCGATATCAAGCATCATAGACCG-3'
Int9NeoF	5'-GTCTCAGAAGGTGGCTGAC-3'
Int9NeoR	5'-CCAGCATTGACCATAGCTGC-3'
Ex1TagF	5'-GCGAGCGCTTTCGGTTTAAC-3'
Ex1TagR	5'-CTCCAGAAGCAGCGGTAGCAAC-3'
Ex10TagF	5'-GTACACCTCACCCATCACAGG-3'
Ex10TagR	5'-CAAACCCATCCAGCCTCCAGTG-3'
3'UpHAF	5'-AAGCTTGCAGGCTGGGTATGGGTATG-3'
3'UpHAR	5'-GGATCCGACCCTATGCAGGTTCTACC-3'
3'DownHAF	5'-GGATCCGAGAAATGCCCCTGGAGACTCTTG-3'
3'DownHAR	5'-GAATTCGAATGTGGCCCCTAGAGATGTC-3'
3'ProbeF	5'-GGGGAAGCTTGACTTGTACCTC-3'
3'ProbeR	5'-CCTTTCCTGTTCTAGCAGTG-3'

2.1.2 Retrieval of the 3' Trf1 knock-in construct for ES cell targeting

To target the 3' *Trf1* knock-in construct to the ES cell genome, the *Trf1* region containing the Venus Tag and neo^r gene needs to be removed from the BAC into an ampicillin resistant retrieval vector. This retrieval vector is constructed with ~500bp *Trf1* homology regions located at ~9kb upstream (5' arm) and ~10kb (3' arm) downstream of the Neo^r insertion site. This will allow for recombination and incorporation of the 3' *TRF1* knock-in construct into the retrieval vector for subsequent ES cell electroporation. The homology regions were PCR amplified from ES cell genomic DNA using the primers 3'UpHAF with 3'UpHAR and 3'DownHAF with 3'Down HAR according to section 2.2 program 2. Approximately 30ng of PCR amplified DNA from each homology region was ligated into ~10ng of pBluescript vector using T4 ligase with the supplied buffer (Promega) and incubated at 4˚C overnight. After incubation, half the ligation mixture was transformed into 75µl DH5α competent *E. coli* bacteria by incubating on ice for 30min, heat-shocking at 42˚C for 90sec, followed by 2min on ice, and incubated in 400µl LB for 1hr at 35˚C. After recovery, 100µl of the transformation was spread on LB agar plates with 20µg/ml ampicillin and incubated overnight at 37˚C. The following day 12 colonies were picked and retrieval vectors (pBluescript with homology regions inserted) were isolated according to the BAC isolation protocol described in section 2.4. The retrieval vectors were digested with 10 Units (U) *Xba*I and 10U *Not*I for verification. A colony with the expected digestion pattern was selected to retrieve the 3' *Trf1* knock-in construct using recombineering as described in section 2.2. Following recombineering 100µl of the bacteria was plated on LB with 10µg/ml kanamycin and incubated at 32˚C for 2 days. Twelve colonies of each of the 5' and 3' constructs were picked and retrieval

vectors were isolated according to section 2.2 and used for restriction digest analysis using 10U *Eco*RV. A positive retrieval vector with the incorporated 3' Trf1 knock-in construct is selected for targeting the ES cell genome by electroporation as described in section 2.5.

2.2 PCR amplification and purification

All PCR mixtures were 25µl and composed of: ~100ng template DNA, 2.5µl of 10X PCR buffer (5ml 1M Tris-Cl pH8.3, 4.9ml dH_2O, 50µl NP-40, 50µl Tween-20), 1µl of 10µM forward and reverse primers (Table 1), 0.75µl of 50mM $MgCl_2$, 0.2µl of 100mM dNTPs, 18µl of dH_2O and 0.5µl of Taq polymerase. The PCR program used was either program 1: 4min at 94˚C, 8X (40sec at 94˚C, 40sec at 56˚C, 2min at 72˚C), 22X (40sec at 94˚C, 40sec at 70˚C, 2min at 72˚C), 7min at 72˚C, followed by 4˚C storage or program 2: 4min at 94˚C, 35X (40sec at 94˚C, 40sec at 54˚C, 2min at 72˚C), 7min at 72˚C, followed by 4˚C storage. PCR products were run on a 1% agarose gel for 1hr at 80V followed by gel purification (Qiagen MinElute gel extraction kit).

2.3 Recombineering

Recombineering was performed according to protocols by Warming et al (2005) using the *galK* selection system.[41] It was performed as follows: 480µl of an overnight culture was added to 24ml of LB medium with chloramphenicol selection (12.5µg/ml) and grown at 32˚C in a shaking waterbath for 2.5 hours. Following this incubation, 12ml was transferred to another 50ml Erlenmeyer flask and heat-shocked at 42˚C for 15min in a shaking waterbath. The remaining culture was left at 32˚C as the uninduced control.

After 15min both samples were cooled on ice, transferred to 15ml Falcon tubes and centrifuged for 3600rpm for 5min at 2˚C. The pellets were washed 3 times by resuspending in 1ml ice-cold ddH_2O by gently swirling the tubes in an ice/water slurry, followed by the addition of 9ml ddH_2O. After the third centrifugation, the supernatant was removed and the pellet was kept on ice until electroporation with 200ng DNA. A 35µl aliquot of bacteria was electroporated in a 0.1cm cuvette at 25µF, 1.75kV, and 200Ω. Electroporated bacteria were then recovered in 1ml LB for 1hr in a 32˚C shaking waterbath. Following recovery, the bacteria were washed twice in 1ml of autoclaved M9 salts (1L: 6g Na_2HPO_4, 3g KH_2PO_4, 1g NH_4Cl, 0.5g NaCl) and pelleted at 13 200rpm for 15sec. After washing, the pellet was resuspended in 1ml of 1X M9 salts and plated according to the *Trf1* knock-in strategy for BAC recombineering.

2.4 BAC isolation

Single bacterial colonies were grown in 5ml of liquid LB medium with 12.5µg/ml of chloramphenicol overnight in a 30˚C shaking incubator. Following incubation, cultures were centrifuged at 3600rpm for 5min. Pellets were then resuspended in 260µl of buffer P1 (50mM Tris pH 8, 10mM EDTA, 100µg/ml RNase A) followed by the addition of 300µl of buffer P2 (0.2N NaOH, 1%SDS). After mixing, the samples were kept at room temperature for 5min after which 300µl of buffer P3 (3M KOAc pH5.5) was mixed in. Samples were then kept on ice for 5min followed by centrifugation at 13,000rpm for 10min. The supernatant was then removed and mixed with 600µl isopropanol and spun at 13,000rpm for 10min to pellet the BAC DNA. After washing the pellet with 500µl of

70% ethanol and spinning at 13,000rpm for 5min, pellets were dried at room temperature and the DNA was resuspended in 30µl nuclease-free water.

2.5 ES cell culture, electroporation and selection

Wild type ES cells from the 129 mouse strain (P14) were cultured on plates coated with 0.1% gelatin in ES medium (Dulbecco's Modified Eagle Medium (DMEM) supplemented with 20% fetal calf serum (FCS), 1X non-essential amino acids, 1X sodium pyruvate, 2mM glutamine, 1X penicillin and streptomycin, and 1000U/ml leukemia inhibitory factor (LIF)). To electroporate the 3' *Trf1* knock-in retrieval vector an ~80% confluent 10cm culture dish was washed in PBS and incubated for 5min at 37˚C in 1ml trypsin. The cells were then pelleted by centrifugation at 1000rpm for 5min in 10ml PBS with 100µl FCS to block the trypsin. Cells were then resuspended in 500µl ES media with 20-40µg DNA and electroporated in a 0.4cm cuvette at 500µF, 240V, and ∞Ω. Following electroporation cells were allowed to recover for 10min in the cuvette at room temperature before being plated in a 10cm culture dish with 10ml ES media and incubated at 37˚C with 5% CO_2.

Electroporated cells were incubated overnight and then placed under selection in ES media with 200µg/ml geneticin for 7 days in either bulk or serial dilutions of 1:10, 1:100, 1:1000, and 1:10000. After 7 days the bulk cultures were selected by fluorescence activated cell sorting (FACS) using the BD FACSVantage SE (Diva Option) based on Venus expression with wild type ES cells as a negative control and ES cells over-expressing Venus-*Trf1* as a positive control (cells from Kathleen Lisaingo). Cells were prepared for sorting by washing with PBS and trypsinizing in 1ml for 5min at 37˚C.

Following this, cells were pelleted in 10ml PBS with 100µl FCS by centrifuging at 1000rpm for 5min. The pellet was resuspended in 1ml PBS with 10% FCS and kept on ice until sorting. Meanwhile, after the 7 days selection, colonies were picked from the serial dilution plates and expanded for DNA extraction (section 2.6) and analysis by Southern blot (section 2.7) to identify colonies where the 3' *Trf1* knock-in construct is correctly targeted to the ES cell genome.

2.6 ES cell DNA extraction

To extract DNA from ES cells, cells were incubated overnight in 500µl lysis buffer (100mM Tris, 5mM EDTA, 0.2% SDS, 200mM NaCl, and 0.2mg/ml proteinase K) at 37˚C. The DNA/lysis buffer was then transferred to a tube containing 500µl phenol chloroform and mixed vigorously. The samples were then centrifuged at 13 000rpm for 10min. After spinning, the upper phase was transferred to a new tube containing 300µl isopropanol, mixed, and centrifuged again at 13 000rpm for 10min. The pellet was then air dried before resuspending in 50µl H_2O

2.7 Southern blot

2.7.1 Probe construction

To make the probe to detect properly targeted 3' *Trf1* knock-in ES cells, PCR amplification as described in section 2.2 program 2 was used with the primers 3'ProbeF and 3'ProbeR (representation of *Trf1* probe location see Figure 10). A 300bp fragment of *Trf1* ~9kb 5' of the neo^r insertion site was amplified for use as a probe. To label this probe, 25ng of DNA was diluted in 18.8µl dH_20, incubated at 100˚C for 3min, and put on

ice for 2min. After this 6µl of oligo loading buffer (100µl solution A (1ml of solution O (1.25M Tris-HCl pH8, and 0.125M $MgCl_2$), 18µl β-mercaptoethanol, 5µl each dNTP except dCTP), 250µl Solution B (2M Hepes pH 6.6), and 150µl solution C (50µl random hexamers Pharmacia stock solution in 550µl TE), 1.5µl of 2ng/ml BSA, 2.5µl α-^{32}P dCTP, and 1.2µl Klenow enzyme were added and the probe was incubated at room temperature for ~5hrs. Following the incubation 50µl Oligo stop solution (20mM Tris-HCl pH8, 2mM EDTA, 20mM NaCl, and 0.25% SDS), 30µl 10mg/ml salmon sperm DNA, 25µl 2M sodium acetate, and 465µl 100% ethanol were added to the probe before centrifugation at 10,000rpm for 15sec. The pelleted probe was washed with 80% ethanol resuspended in 500µl dH_20, denatured at 100˚C for 3min, and stored on ice until used for Southern analysis.

2.7.2 Southern blot analysis

Southern blots were performed by digesting the genomic ES cell DNA with 25U *Hpa*I overnight at 37˚C. After digestion, 25µl of the DNA was run on a 0.8% agarose gel at 60V overnight. The gel was then processed for blotting as follows: 10min wash with depurination solution (0.25M HCl), 20 min wash with denaturation solution (1.5M NaCl, 0.5M NaOH), and 30 min wash with neutralization solution (1L: 87.66g NaCl, 60.5g Tris pH 7.5) rinsing with water in between treatments. Blotting transfer to Hybond N^+ membrane (Amersham) was then set up overnight, and the following day the DNA was fixed to the membrane by washing with 0.4M NaOH for 10min. Prehybridization was performed by incubating the membranes in 25ml hybridization solution (1L: 500ml phosphate buffer pH 7.2 (342ml 1M Na_2HPO_4 and 158ml 1M NaH_2PO_4), 350ml 20%

SDS, 2ml 0.5M EDTA, 148ml H_2O) and incubation at 62˚C for 1hr in a rotating hybridization oven. Hybridization was then carried out by replacing the solution with 25ml hybridization solution containing the labeled probe and incubated at 62˚C overnight. Following a 2X 15min wash with 2X SSC (20X(1L): 88.23g Tri-sodium citrate, 175.32g NaCl pH 7) 0.1% SDS and a 2X 15min wash with 1X SSC 0.1% SDS, blots were placed on a phosphorimager cassette at room temperature overnight and analyzed using a phosphorimager (STORM 860 molecular imager) the following day.

Chapter 3: Results

3.1 Development of the *Trf1* knock-in constructs

The strategy used when building the 5' knock- in construct was to insert a Tag into exon 1 of the *Mus musculus* gene *Trf1*directly 3' of the ATG start codon (Figure 5). This Tag consists of a Venus fluorescent marker that can be used for microscopic visualization of TRF1 at the telomeres. This will allow visualization and tracking of telomeres using live cell imaging. A 6X His and a 3X FLAG peptide were also incorporated into the Tag for immunoprecipitation that can be used to identify TRF1-associating proteins. A TEV cut site is included between the FLAG-Venus and 6X His so the tag can be cleaved to reduce bulkiness or to allow an additional immunoprecipitation step before and after cleavage. The Tag is separated from the TRF1 protein by a flexible linker to prevent the Tag from disrupting TRF1 function or telomere binding. In addition, a neomycin resistance (neo^r) cassette was targeted to intron 1 (Figure 5). This location was chosen because genomic integration events can be selected for once they are targeted into the mouse ES cells but translation of the tagged TRF1 will not be disrupted. The restriction enzyme cut site *Avr*II was engineered into the neo^r cassette so that Southern blotting can be used to identify properly targeted ES cells. In properly targeted cells, genomic DNA digested with *Avr*II will produce a smaller band due to the inserted additional cut site between the two flanking endogenous cut sites that are external to the homology arms.

The strategy used to build the 3' knock- in construct was similar to that of the 5' construct. In this case however, a Tag with identical components to those described

above was targeted to exon 10 directly 5' of the TGA stop codon (Figure 5). The neo^r cassette was targeted to intron 9 to avoid disrupting protein translation. Within the neo^r cassette the restriction enzyme cut site, *Hpa*I, was added for Southern blot analysis to identify properly targeted ES cells.

The first recombineering step to develop these constructs was to target neo^r to the desired insertion sites. To do this a neo^r PCR product, engineered with ~50-70bp homology arms corresponding to the desired *Trf1* insertion site, was electroporated into bacteria harboring a BAC containing the *Trf1* gene (Trf1/BAC) (Figure 6, Step 1). In bacteria the neo^r gene confers resistance to kanamycin, therefore to screen for proper recombination events, electroporated bacteria were plated on kanamycin-containing media. BACs were purified from the kanamycin-selected colonies and the putative insert regions were amplified by PCR. Positive colonies (Trf1/BAC-Neo) are characterized by a ~1800bp increase in size compared to a control PCR of the Trf1/BAC without the neo^r (Figure 7). The *galK* gene was then inserted into Trf1/BAC-Neo by recombination using homology arms targeting to the desired location for the Tag (Figure 6, Step 2). This provides the selection step where colonies containing BACs that have swapped the tag in place of *galK* will grow in the presence of DOG (see Materials and Methods 2.1.1 for details). To verify correct targeting of *galK* in Trf1/BAC-Neo, PCR amplification across the insert site was performed as described with neo^r insertion. A size increase of ~1200bp over the control indicates proper insertion (Figure 7).

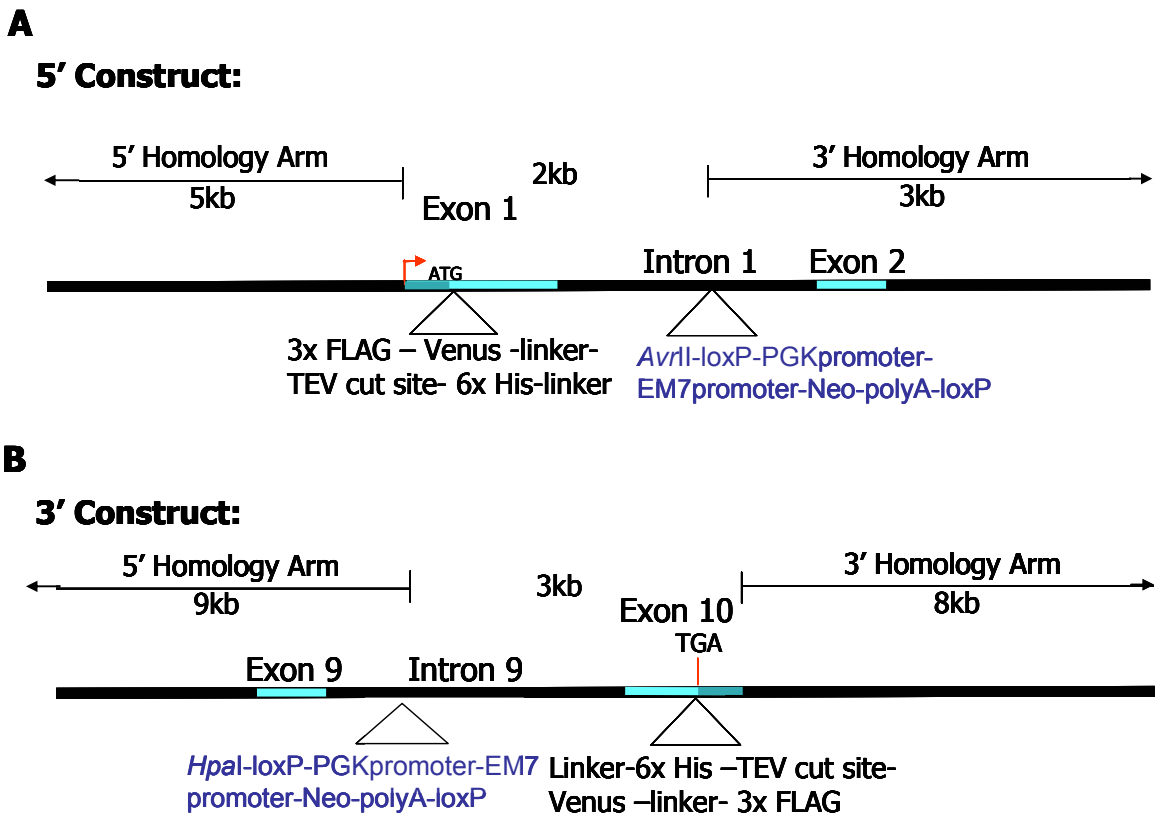

Figure 5. Representation of the *Trf1* knock-in constructs produced through recombineering. A. The 5' construct includes a Tag inserted directly 3' of the ATG start codon in exon 1. It consists of a Venus fluorescent protein for visualization of TRF1 telomere association and both a 3X FLAG and 6XHis, separated by a TEV cut site, for immunoprecipitation. A linker separates the Tag region from TRF1 to prevent disrupting the protein function. For selection, the construct also contains a neo^r cassette targeted to intron 1 with the addition of the *Avr*II cut site that will be used for Southern blot analysis to identify properly targeted ES cells. The insert regions are flanked by long *Trf1* homology arms for recombination into the ES cell genome. B. The 3' construct has a similar Tag targeted to exon 10 directly 5' of the TGA stop codon. The neo^r cassette was inserted in intron 9 with the addition of the *Hpa*I cut site.

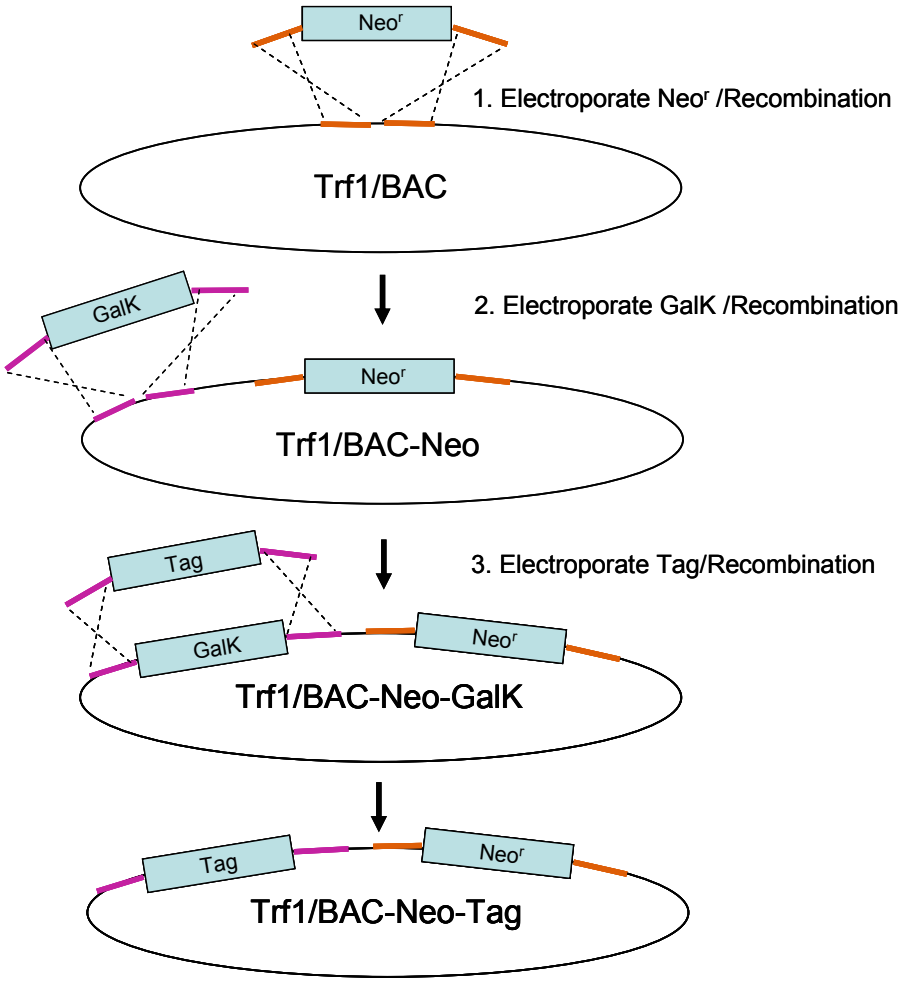

Figure 6. Representation of the BAC recombineering steps involved in assembling the *Trf1* knock-in constructs. The first step was to electroporate the PCR amplified neo^r cassette, engineered with homology arms corresponding to the desired insertion site, into bacteria with Trf1/BAC. Recombination incorporated the neo^r to produce Trf1/BAC-Neo. The next step was electroporating PCR amplified *galK,* with homology arms targeted to the insertion site of choice, into bacteria containing Trf1/BAC-Neo for recombination producing Trf1/BAC-Neo-GalK. The final step was to electroporate the Tag, with identical homology arms to *galK,* into bacteria with Trf1/BAC-Neo-GalK for recombination with *galK* producing the final knock-in constructs within the BACs (Trf1/BAC-Neo-Tag).

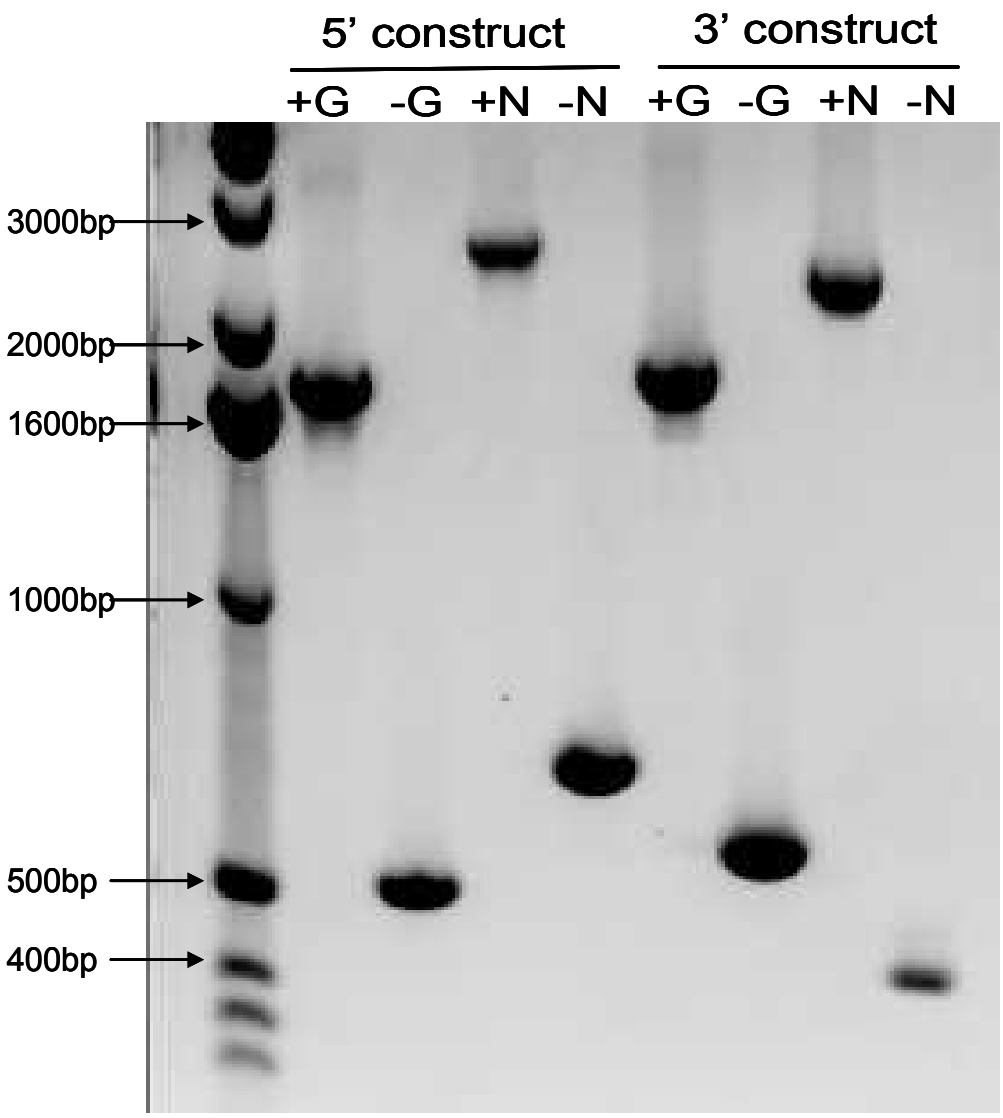

Figure 7. Agarose gel showing the PCR reaction products verifying properly targeted *galK* and neo^r in the *Trf1* knock-in constructs. PCR amplification was performed across the insert region for both *galK* (G) and neo^r (N) in both the 5' and 3' constructs. Introduction of the respective cassettes is shown by a size increase of ~1800bp for neo^r insertion (+N) and ~1200bp for *galK* insertion (G+) from the Trf1/BAC control (-N, -G) PCR.

The final step in developing the TRF1 knock-in constructs was to replace the *galK* cassette with a cassette containing the amplified Tags with the same *Trf1* homology arms as *galK* (Figure 6, Step 3). The Tag was electroporated into Trf1/BAC-Neo-GalK for recombination to occur. Colonies from this electroporation were grown on DOG minimal media plates for selection of properly targeted Tags. To verify that the Tag had properly recombined into the construct, Trf1/BAC-Neo-Tags were purifed from selected colonies and PCR amplified across the putative target site. To ensure that the Tag is targeted and free from mutations that could occur during this process, the PCR product was sequence analyzed. Cultures with proper BAC *Trf1* knock-in constructs were used in subsequent experiments.

3.2 Testing functionality of the tag in the *Trf1* knock-in constructs

To quickly test that the tags of the *Trf1* knock-in constructs are functional, Trf1/BAC-Neo-Tags for both the 5' and 3' constructs were isolated from bacteria and electroporated into wild-type ES cells to transiently over-express the tagged protein. This quick test to verify construct functionality is possible because the *Trf1* promoter, along with the entire gene sequence and inserted tag, is present within the BAC. The electroporated ES cells were sorted based on Venus fluorescence (Figure 8). Positive cells were gated using a negative control (wild-type ES cells) and a positive control (ES cells over-expressing Venus-TRF1). Both the 5' and 3' constructs produced cells expressing Venus fluorescence indicating that the Tag within the TRF1 knock-in constructs produces fluorescence and should be retrieved from the BACs for ES cell genome targeting.

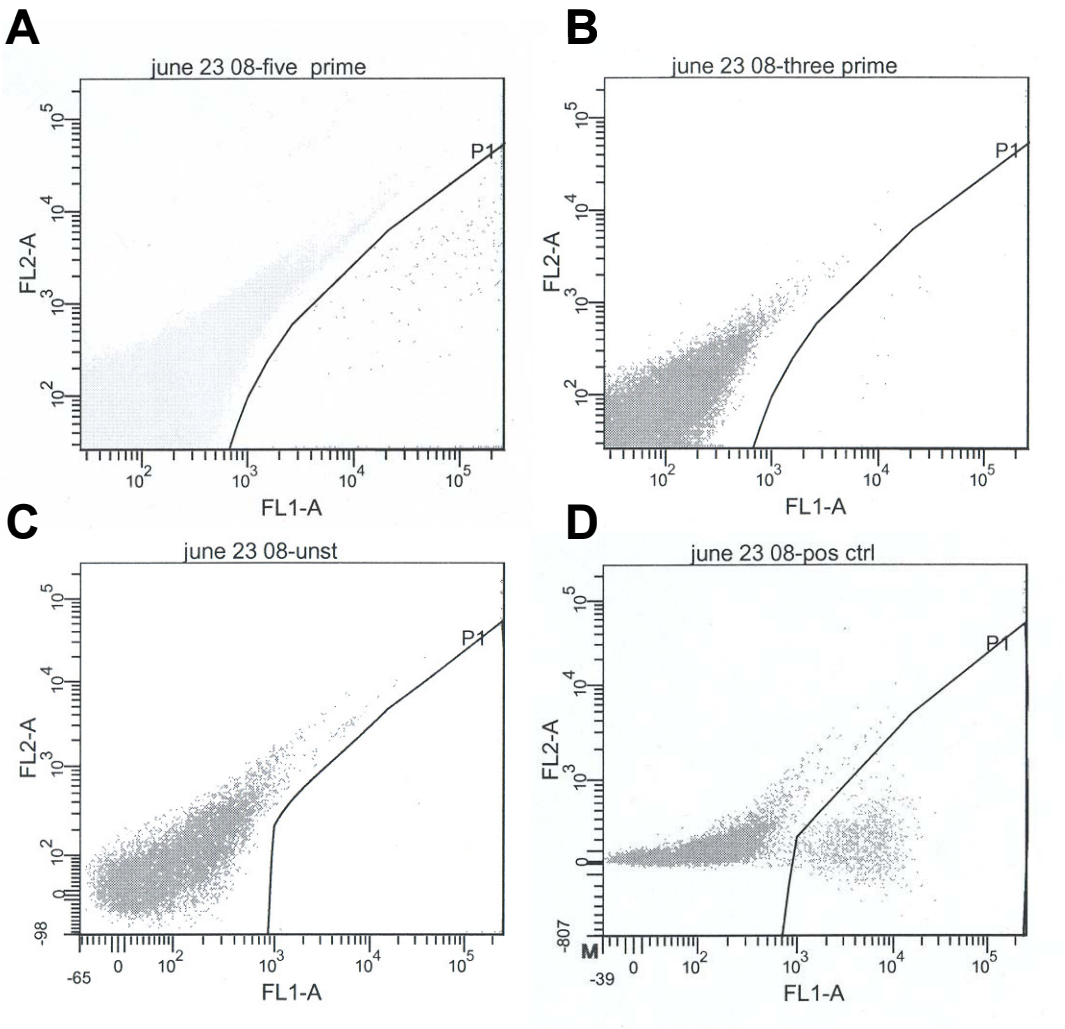

Figure 8. Flow-sorting of ES cells electroporated with Trf1/BAC-Neo-Tag based on Venus expression. Each dot on the plot represents an individual cell and Gate P1 was drawn to select and sort Venus positive cells. A. Electroporation of the 5' *Trf1* knock-in construct in Trf1/BAC-Neo-Tag produced 0.018% of cells expressing Venus fluorescence. B. Electroporation of the 3' *Trf1* knock-in construct in Trf1/BAC-Neo-Tag produced 0.034% of cells expressing Venus fluorescence. C. Negative control using wild type ES cells and D. Positive control of ES cells electroporated with a Venus-TRF1 over-expression vector allow for the selection of Gate P1 to sort Venus positive cells.

3.3 Targeting the 3' *Trf1* knock-in construct to the mouse ES cell genome

To target the 3' *Trf1* knock-in construct to the ES cell genome, a retrieval vector (pRV) is used to isolate the construct from Trf1/BAC-Neo-Tag (Figure 9) This retrieval vector is constructed by cloning 300-500bp PCR-purified *Trf1* homology regions into a pBluescript backbone. The homologous regions were chosen to be ~9kb upstream and downstream of the neo^r insert so that recombination between Trf1/BAC-Neo-Tag and pRV will result in a retrieval vector containing the construct flanked by these ~9kb long homology arms for ES cell genome targeting (pRV-5'/3'Con). This plasmid, following restriction digest verification with *Eco*RV, was purified in preparation of electroporation into ES cells.

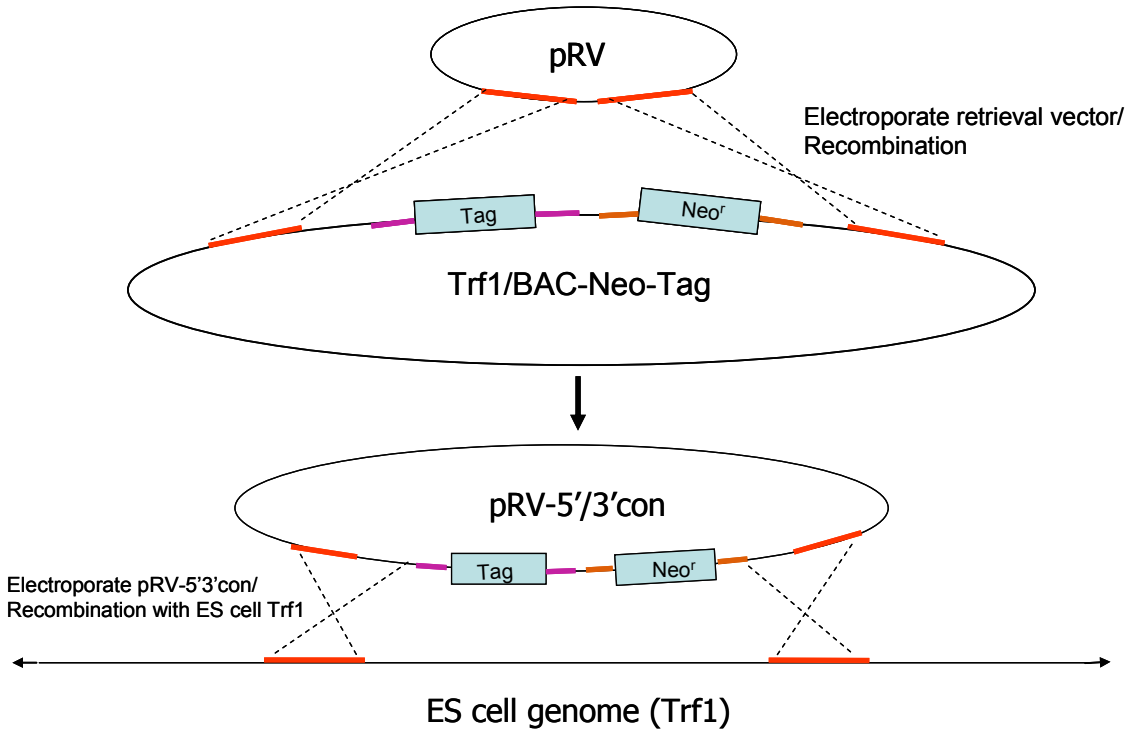

Figure 9. Representation of the steps involved in retrieving the *Trf1* knock-in construct from BACs and targeting it to the ES cell genome. The retrieval vector pRV containing *Trf1* homology regions flanking the knock-in construct was electroporated into bacteria containing Trf1/BAC-Neo-Tag. Recombination results in the incorporation of the knock-in construct into pRV producing pRV-5'/3'con which was then electroporated into ES cells for targeting and recombination into the *Trf1* gene within the ES cell genome.

After electroporation of pRV-3'con into wild-type ES cells and selection for seven days with geneticin, colonies were picked and expanded to allow for enough cells to be harvested for DNA extraction and Southern blot analysis to verify *Trf1* targeting. By digesting the DNA with *Hpa*I, the additional cut site inserted with the neo^r cassette will reduce the probed fragment size from 19kb to 9kb as long as the construct was properly targeted to the *Trf1* gene. This reduction in size of the probed fragment will allow for easy identification of positively targeted constructs by producing a 19kb and a 9kb band in the Southern blot analysis. Southern blots were performed on all 78 ES cell colonies that were harvested post-electroporation with the TRF1 knock-in construct and post-selection with geneticin. All samples in the analysis showed a band at 19kb while none displayed an additional 9kb band (Figure 10). This suggests that none of the selected colonies were correctly targeted and more colonies need to be tested.

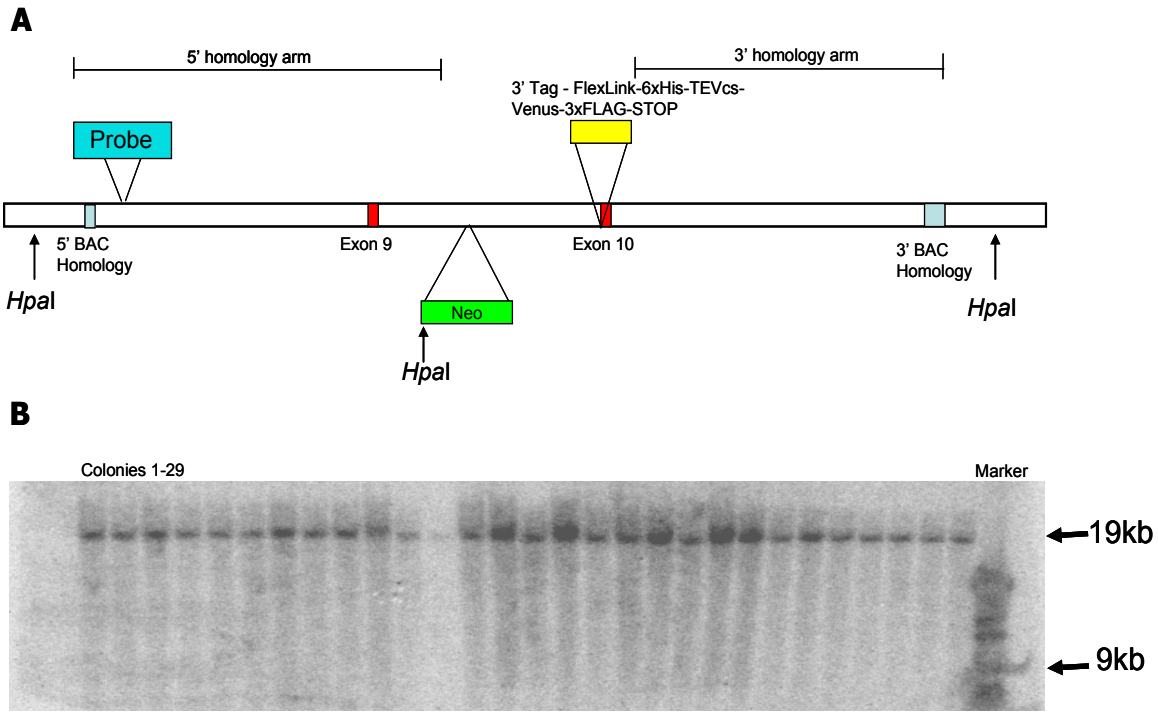

Figure 10. Southern blot of ES cell colonies electroporated with the 3' *Trf1* knock-in construct in pRV-3'con. A. Representation of the 3' *Trf1* knock-in construct in the target region. The endogenous *Hpa*I cut sites are 19kb apart. If the *Trf1* locus was properly targeted, the *Hpa*I cut site inserted with neo^r will reduce the probed region to 9kb. B. Southern blot analysis where *Hpa*I digested ES cell genomic DNA was run on an agarose gel before transfer to a Hybond N+ membrane and probed with a PCR generated probe located 5' of the knock-in construct insertion site. Phosphoimager detection was used to visualize probed DNA. A *Trf1* locus that was properly targeted should have both a 19kb and a 9kb band. The absence of a 9kb band suggests that no ES cells incorporated the construct into the right locus.

3.4 Cell sorting of Venus expression from the 3' *Trf1* knock-in construct

While individual colonies did not yield a positively targeted 3' construct, ES cells were also electroporated with pRV-3'con and bulk-sorted based on Venus fluorescence (Figure 11). Because TRF1 is not a highly expressed protein, the level of fluorescence in successfully electroporated cells is not expected to be very high relative to control cells. Therefore, cells were gated relatively close to the negative population and relative to the wild-type ES cell negative control. The sort shows approximately a 0.01% proportion of fluorescent cells indicating that approximately 1 in 100 cells are expressing the inserted Tag. While this does not necessarily demonstrate that the insert in these cells is correctly targeted to the *Trf1* gene, it does suggest that well over 100 colonies may have to be screened to find a positive.

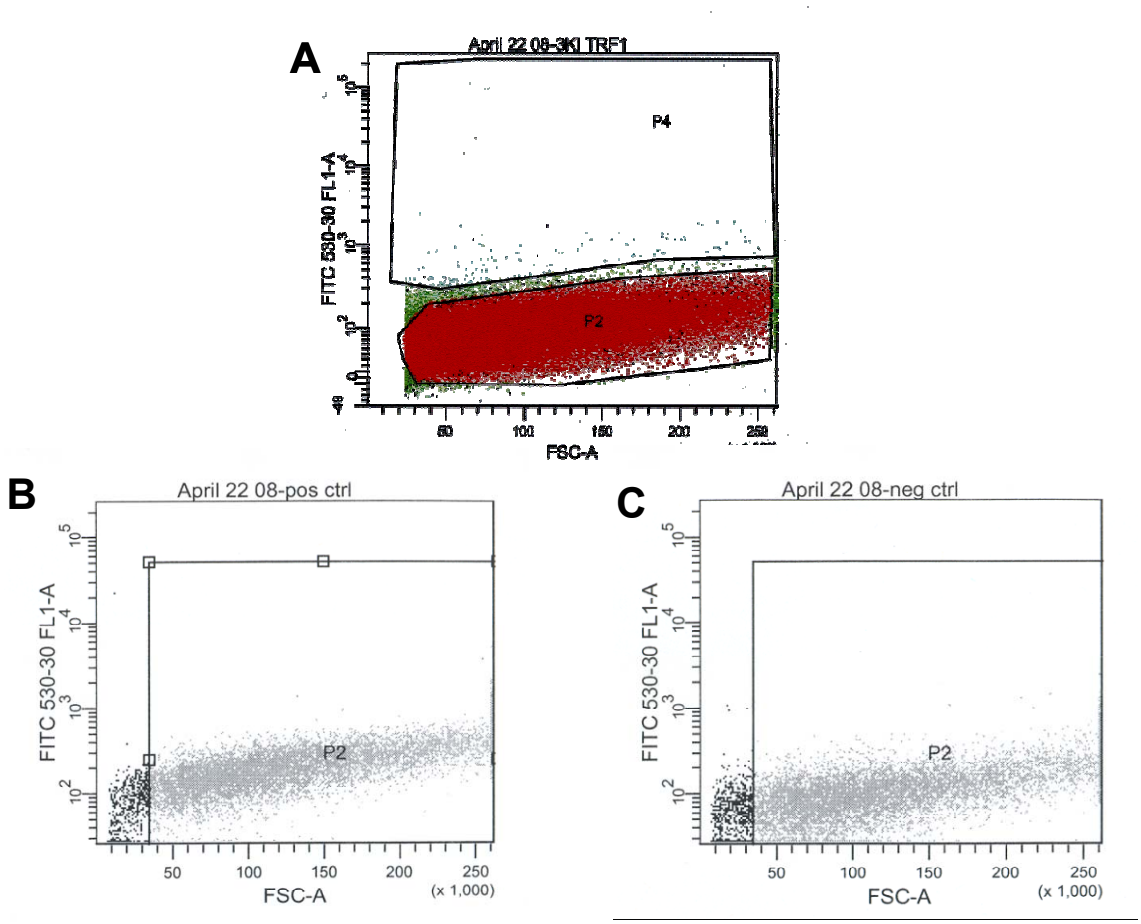

Figure 11. Cell sorting based on Venus expression of bulk ES cells electroporated with the 3' *Trf1* knock-in construct. A. Each dot on the plot represents an individual cell and Gate P4 was drawn to select and sort Venus positive cells (green) based on: B. A positive control sort of knock-in ES cells expressing Venus-*Rtel* from the endogenous promoter (another very low expressing protein) and C. A negative control sort of wild-type ES cells. The sort produced a 0.01% population of positive cells.

The successful sorting of Venus expressing cells from the electroporation of pRV-3'con indicates that the Venus Tag is functional within the construct. To generate a properly-targeted *Trf1* knock-in cell line, additional electroporations and colony screening will be performed. The 5' *Trf1* knock-in construct currently remains within BACs. After the retrieval of the construct using pRV, pRV-5'con will be electroporated into ES cells to generate *Trf1* knock-in cell lines.

Chapter 4: Discussion

Telomeres are important biological structures to study because they play a critical role in many disorders as well as normal biological processes. Using TRF1 as a tool to examine telomere dynamics is practical due to its specificity for telomeric repeats and its ubiquitous expression.

The goal of this project was to develop TRF1 knock-in ES cells using recombineering in order to solve problems associated with TRF1 over-expression such as accelerated telomere shortening. While this method has a number of advantages over traditional genetic engineering such as not being limited by restriction enzyme cut sites and the larger size of DNA fragments that can be incorporated, there were a variety of challenges that required troubleshooting along the way. These include relatively simple electroporation variables such as keeping cells competent and determining and using appropriate DNA and cell concentrations. Another problem that can be encountered during this process is internal BAC recombination. This unpreventable problem can occur at any time during BAC recombineering. Its frequency is dependent on the sequence content of the individual BAC being used. Therefore, the many steps to verify the integrity of BACs at each step by restriction digest analysis throughout the process are essential to achieve successful recombineering.

The region containing the complete 3' *Trf1* knock-in construct was successfully retrieved from the BAC and electroporated into ES cells (Figure 6 Step 4). Of the 78 colonies that grew after selection, none of them showed proper targeting as determined by Southern analysis. When constructs were electroporated and bulk-sorted based on

Venus fluorescence, only 0.01% of the cells showed fluorescence. Since it is expected some constructs will not be correctly targeted and may insert near a promoter, this could mean that hundreds of colonies may have to be analyzed to find a single positive colony. Other factors can also affect targeting of the construct into the ES cell genome. For example, it is possible that the neo^r was recombined into the ES cell genome without the Tag component. This will produce a number of false positives after selection. In addition, the targeting vector pRV-5'/3'con can form secondary structure reducing its recombination efficiency. Lastly, the ES cell chromatin structure where the *Trf1* knock-in construct is targeted can affect the efficiency of recombination. To overcome factors such as these, the efficiency of targeting needs to be improved from the levels obtained in the present study.

Targeting efficiency of the knock-in construct to the ES cell genome can be improved in a number of ways. First, the concentration of geneticin and length of time under selection can be increased. This may reduce the number of false positives when screening the colonies by Southern blot analysis. Another factor that can be optimized for increased targeting efficiency is the concentration of DNA used in the electroporation. If the secondary structure of the vector or chromatin structure in the target region is making recombination difficult, the amount of electroporated DNA may need to be increased to compensate for these challenges. In addition, the length of homology arms for targeting can be adjusted in an effort to improve targeting efficiency. Again, if the chromatin structure in the target region is tightly packaged, shorter homology arms may be necessary to allow invasion of this complex structure and allow recombination to occur. Lastly, the sequence of the homology arms needs to be taken into consideration. The *Trf1*

gene in the BAC is from the mouse strain Black/6 and is being targeted to the ES cell genome of the 129 mouse strain. The sequence between these two strains should be analyzed to verify that it is sufficiently homologous for recombination to occur efficiently. If the sequence between these two strains differs, the probability of proper genome targeting could be considerably decreased.

Once a colony is identified that has the knock-in construct properly targeted to *Trf1* within the ES cell genome, it is important to ensure endogenous levels of expression. One factor that could affect expression levels is the addition of neo^r within an intron of *Trf1*. This insertion could result in aberrant expression by affecting the efficiency of association of transcription machinery or by producing incomplete transcription products. There is a possibility that transcription could begin at the endogenous *Trf1* promoter and terminate prematurely at the neo^r transcription terminus. This can be avoided by removing the neo^r cassette once a properly targeted colony is identified. The addition of LoxP sites flanking the cassette allows for removal through the Lox-Cre recombination system. Once removed, the resulting cell line should be expanded and analyzed by western blot to ensure that endogenous levels of TRF1 are obtained.

The question remains as to which of the two constructs retains endogenous functionality. The 5' construct will result in the tag being incorporated at the N-terminus of the TRF1 protein next to the regulatory acidic domain [9]. This is located on the opposite end of the DNA binding domain with the TRF specific domain for homodimerization between the tag and DNA binding domain (Figure 2). Over-expression constructs described within the literature and made within the lab fused the tag to the N-terminus of the TRF1 protein [31]. Deletions along the N-terminus up to and including some of the

acidic regulatory domain have not disrupted protein binding suggesting that this terminus is not critical for protein localization to the telomere [43]. This provides strong support for the construction of a 5' *Trf1* knock-in ES cell line to examine telomere dynamics.

A 3' *Trf1* knock-in construct was also developed despite the absence of 3' TRF1 fusions in the literature. The TRF1 protein contains a C-terminal DNA binding domain that targets the protein to telomeres. When making the 3' *Trf1* knock-in construct the assumption is made that the addition of the Tag will not disrupt DNA binding, however this needs to be tested. The addition of the flexible linker between the Tag and TRF1 protein will ideally ensure that the protein binding ability will be maintained in *Trf1* knock-in cell lines.

The development of these *Trf1* knock-in ES cell lines will allow for the dynamics of telomeres to be studied under a variety of conditions. This tool will allow telomere dynamics to be examined within ES cells during cell division, differentiation, knock-down of other proteins, and in response to stress conditions. For example, differentiating *Trf1* knock-in ES cell lines can be used to visually examine properties such as telomere length and movement during the differentiation process as well as in somatic cells throughout their proliferative life span and into senescence. As discussed previously, this would be useful to track changes to telomeres during the "aging" process. In addition, if a protein of interest is thought to affect telomere function, the expression of the protein could be knocked-down in *Trf1* knock-in ES cells and telomeres could be examined for disruptions in functions such as chromosome end protection. This tool is also valuable in examining telomere response to stress. For example, by exposing the cells to different stressors such as oxidative stress, chemotherapeutic agents, or other drugs, effects on

telomeres can be visually examined. If the stressor causes disruption in telomere function, we may observe genomic instability similar to that seen in cancer or other genomic instability disorders.

In addition, advances in selecting cell types by *in vitro* differentiation will allow for telomeres to be examined in specific cells. The FLAG peptide can be used to identify novel proteins that may associate with TRF1 in different conditions. For example, it is possible to differentiate ES cells to isolate primordial germ cells [44]. With this cell line telomere dynamics could be tracked for the first time using live-cell imaging during the specialized function of bouquet formation during gametogenesis. Specialized proteins that may be recruited and interact with TRF1 during this process could be discovered using the FLAG tag. The knock-down of these novel binding partners could demonstrate that they are required for bouquet formation and chromosome segregation.

These *Trf1* knock-in ES cell lines could be used *in vitro* to answer a number of questions regarding telomere dynamics, some ideas of which have been discussed here. A longer term goal for these constructs is to generate TRF1 knock-in mice for *in vivo* studies of telomere dynamics. These mice will allow for studies of telomere dynamics and novel TRF1 binding proteins in different tissues and during different stages of development. For example, *in vivo* gametogenesis could be examined through tubule squashes to see if observations of bouquet formation and chromosome segregation in culture are also observed within the animal. In addition, telomere dynamics can be observed in conditions such as chemotherapy treatment to see how their functionality is affected. Lastly, somatic cells from animals with a condition such as an accelerated aging

syndrome or a genomic instability disorder can be compared to wild type tissue for disruptions in telomere location, length, function, or protein associations.

Development of this tool is important to examine telomeres in different experimental conditions without the problems associated with TRF1 over-expression. By adding a tag to the *Trf1* gene that is regulated at endogenous levels rather than a protein over-expression system, we can be certain that our observations are the best representation of normal telomere dynamics. The wide range of telomere functions and associations results in the need for a TRF1 knock-in ES cell line to accurately examine telomeres in a variety of experimental conditions.

References

1. Aubert, G. & Lansdorp, P. M. Telomeres and aging. *Physiol. Rev.* **88**, 557-579 (2008).

2. Lansdorp, P. M. Major cutbacks at chromosome ends. *Trends Biochem. Sci.* **30**, 388-395 (2005).

3. McEachern, M. J., Krauskopf, A. & Blackburn, E. H. Telomeres and their control. *Annu. Rev. Genet.* **34**, 331-358 (2000).

4. Liu, D., O'Connor, M. S., Qin, J. & Songyang, Z. Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *J. Biol. Chem.* **279**, 51338-51342 (2004).

5. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100-2110 (2005).

6. Zhong, Z., Shiue, L., Kaplan, S. & de Lange, T. A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Mol. Cell. Biol.* **12**, 4834-4843 (1992).

7. Kanoh, J. & Ishikawa, F. Composition and conservation of the telomeric complex. *Cell Mol. Life Sci.* **60**, 2295-2302 (2003).

8. Bilaud, T. *et al*. Telomeric localization of TRF2, a novel human telobox protein. *Nat. Genet.* **17**, 236-239 (1997).

9. Ye, J. Z. *et al*. TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J. Biol. Chem.* **279**, 47264-47271 (2004).

10. Baumann, P. Are mouse telomeres going to pot? *Cell* **126**, 33-36 (2006).

11. Li, B., Oestreich, S. & de Lange, T. Identification of human Rap1: implications for telomere evolution. *Cell* **101**, 471-483 (2000).

12. Smogorzewska, A. *et al*. Control of human telomere length by TRF1 and TRF2. *Mol. Cell. Biol.* **20**, 1659-1668 (2000).

13. Griffith, J. D. *et al*. Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-514 (1999).

14. Zakian, V. A. Structure and function of telomeres. *Annu. Rev. Genet.* **23**, 579-604 (1989).

15. Wai, L. K. Telomeres, telomerase, and tumorigenesis--a review. *MedGenMed* **6**, 19 (2004).

16. de Lange, T. Protection of mammalian telomeres. *Oncogene* **21**, 532-540 (2002).

17. Jordan, P. Initiation of homologous chromosome pairing during meiosis. *Biochem. Soc. Trans.* **34**, 545-549 (2006).

18. Chikashige, Y. *et al*. Meiotic proteins bqt1 and bqt2 tether telomeres to form the bouquet arrangement of chromosomes. *Cell* **125**, 59-69 (2006).

19. Keefe, D. L., Marquard, K. & Liu, L. The telomere theory of reproductive senescence in women. *Curr. Opin. Obstet. Gynecol.* **18**, 280-285 (2006).

20. Shay, J. W. & Wright, W. E. Aging. When do telomeres matter? *Science* **291**, 839-840 (2001).

21. Shay, J. W. & Wright, W. E. Telomeres and telomerase: implications for cancer and aging. *Radiat. Res.* **155**, 188-193 (2001).

22. Shin, J. S., Hong, A., Solomon, M. J. & Lee, C. S. The role of telomeres and telomerase in the pathology of human cancer and aging. *Pathology* **38**, 103-113 (2006).

23. Shay, J. W. & Wright, W. E. Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis* **26**, 867-874 (2005).

24. Meeker, A. K. & De Marzo, A. M. Recent advances in telomere biology: implications for human cancer. *Curr. Opin. Oncol.* **16**, 32-38 (2004).

25. Harley, C. B. Telomerase and cancer therapeutics. *Nat. Rev. Cancer.* **8**, 167-179 (2008).

26. Liu, L. *et al*. Irregular telomeres impair meiotic synapsis and recombination in mice. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6496-6501 (2004).

27. Molenaar, C. *et al*. Visualizing telomere dynamics in living mammalian cells using PNA probes. *EMBO J.* **22**, 6631-6641 (2003).

28. Broccoli, D. *et al*. Comparison of the human and mouse genes encoding the telomeric protein, TRF1: chromosomal localization, expression and conserved protein domains. *Hum. Mol. Genet.* **6**, 69-76 (1997).

29. Houghtaling, B. R., Cuttonaro, L., Chang, W. & Smith, S. A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Curr. Biol.* **14**, 1621-1631 (2004).

30. van Steensel, B. & de Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740-743 (1997).

31. Mattern, K. A. *et al*. Dynamics of protein binding to telomeres in living cells: implications for telomere structure and function. *Mol. Cell. Biol.* **24**, 5587-5594 (2004).

32. Smith, S. & de Lange, T. TRF1, a mammalian telomeric protein. *Trends Genet.* **13**, 21-26 (1997).

33. Karlseder, J. *et al*. Targeted deletion reveals an essential function for the telomere length regulator Trf1. *Mol. Cell. Biol.* **23**, 6533-6541 (2003).

34. Chen, Y. *et al*. A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* **319**, 1092-1096 (2008).

35. Bianchi, A. *et al*. TRF1 binds a bipartite telomeric site with extreme spatial flexibility. *EMBO J.* **18**, 5735-5744 (1999).

36. Ohki, R. & Ishikawa, F. Telomere-bound TRF1 and TRF2 stall the replication fork at telomeric repeats. *Nucleic Acids Res.* **32**, 1627-1637 (2004).

37. Hanaoka, S., Nagadoi, A. & Nishimura, Y. Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. *Protein Sci.* **14**, 119-130 (2005).

38. Iwano, T., Tachibana, M., Reth, M. & Shinkai, Y. Importance of TRF1 for functional telomere structure. *J. Biol. Chem.* **279**, 1442-1448 (2004).

39. Loayza, D. & De Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **423**, 1013-1018 (2003).

40. Copeland, N. G., Jenkins, N. A. & Court, D. L. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2**, 769-779 (2001).

41. Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005).

42. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13**, 476-484 (2003).

43. Bianchi, A., Smith, S., Chong, L., Elias, P. & de Lange, T. TRF1 is a dimer and bends telomeric DNA. *EMBO J.* **16**, 1785-1794 (1997).

44. Geijsen, N. *et al*. Derivation of embryonic germ cells and male gametes from embryonic stem cells. *Nature* **427**, 148-154 (2004).