Oncofetal HMGA2 as a key regulator of telomere stability through control of DNA topology

Dharshana, Ramani Priya

2018

Dharshana, R. P. (2018). Oncofetal HMGA2 as a key regulator of telomere stability through control of DNA topology. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/89230>

<https://doi.org/10.32657/10220/46217>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**ONCOFETAL HMGA2 AS A KEY REGULATOR OF
TELOMERE STABILITY THROUGH CONTROL OF
DNA TOPOLOGY**

RAMANI PRIYA DHARSHANA

SCHOOL OF BIOLOGICAL SCIENCES

2018

**ONCOFETAL HMGA2 AS A KEY REGULATOR
OF TELOMERE STABILITY THROUGH
CONTROL OF DNA TOPOLOGY**

RAMANI PRIYA DHARSHANA

School of Biological Sciences

**A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy**

2018

ACKNOWLEDGMENTS

I express my deepest gratitude to my supervisor Prof. Peter Droge for his valuable guidance, encouragement and expert supervision. His tremendous enthusiasm for science has been a source of inspiration throughout my PhD and will continue far into the future. I feel very fortunate to have been mentored by Dr. Droge and will cherish this experience forever.

I gratefully acknowledge Dr. Roland Ivanyi-Nagy for his wholehearted support, critical review, valuable inputs and constructive suggestions. His willingness to help without hesitation requires special mention.

I am grateful to Costerwell Khyriem (NTU) for his immense help in the Bioinformatics analysis and teaching me the basics. I wish to thank Barbara (NTU) for training me to use the confocal microscope.

I wish to convey special thanks to the past members of the Droge lab - Dr. Yu Haojie, for helping me learn the ropes of working in this study, Dr. Shree Harsha Vijaya Chandra, for guiding me in the cell line engineering work.

My appreciation also goes to my labmates, Sabrina Peter, Dr. Bao Qiuye, Syed Moiz and Suki Roy for providing support and a pleasant work environment during my research.

I feel blessed to have Harshyaa Makhija, as both a dear friend and a valuable colleague. She has always been a constant support and has been my family in Singapore. I wish to thank my dear friend Moumita for patiently helping me with the revision. I owe special thanks to Dr. Abishek Sridhar for his valuable inputs on my thesis and Dr. Muneesh Maheswari for helping me out with formatting.

I wish to thank my incredible friends Hema & Bhaskar Peri, Pravesh, Manvi, Divya, Anjana, Saranya, Shuba, Tofayel, Arvind, Manisha, Malini, Tanya and Nitin for their help and moral support during my research.

I am grateful for the opportunity and the financial support by Telomere Dynamics Group (TDG), Nanyang Technological University, Singapore.

Finally, I acknowledge and thank my family. Their unconditional love has always been my moral strength and support. My deepest appreciation goes to my parents and brother, whose encouragement and confidence has helped me survive many difficult moments. I also extend my respect to my parents-in-law, Sasi pedhamma and Kumar mama. My special love goes to my husband, CM, who has been my dearest friend and continues to share both my successes and difficulties. With my family beside me, I am confident that I can reach the greatest heights.

*This thesis is dedicated to the greatest daddy I could
ever ask for*

ABSTRACT

Faithful replication of the whole genome is fundamental to maintain the genome integrity. During this process, the replication forks may encounter various endogenous and exogenous factors that could lead to replication fork stalling. Telomeres in particular experience numerous replication and topological fork barriers due to its intrinsic nature characterized by repetitive DNA sequences, heterochromatin, and formation of DNA secondary structures; a consequence of TERRA transcription. These impediments present telomeres as ‘difficult-to-replicate’ regions in the genome necessitating both temporal and spatial barrier resolution, which could otherwise lead to telomere deprotection and accumulation of DNA aberrations; a hallmark of cancer. To this end, tightly regulated mechanisms exist to ensure high fidelity telomeric replication.

To study the various aspects of telomeres biology, I used the CRISPR/Cas9 tool to generate human embryonic stem cells and HT1080 fibrosarcoma cells stably expressing either GFP-FLAG or mCherry-miniSOG at the C-terminus of the shelterin protein TRF1. TRF1 interacts directly with telomeres and the visualization will help gain insights about telomere regulation in embryonic cells compared to malignant cells. The 3xFLAG tag will enable pull-down studies while the miniSOG will be used to induce telomere site-specific reactive oxygen species (ROS) to study the oxidative stress at telomeres.

Recent studies identified the novel role of the oncofetal, non-histone chromatin protein, High Mobility Group AT-hook 2 (HMGA2) in providing first-line defense mechanism against endonucleolytic collapse of stalled forks in embryonic stem cells and cancer cells (Yu, Lim et al. 2014). However, the precise mechanism by which HMGA2 recognizes and interacts with the replication fork remains unclear.

Herein, we demonstrate that the chaperoning effect of HMGA2 is conserved across a spectrum of cell lines in addition to ESCs and cancer cells. In the context of malignant cells, we discover differential chemosensitivity effect on the HMGA2 expressing cancer cells in response to chemotherapeutic drugs targeting

Topoisomerase I, which is attributed to variable expression levels of HMGA2 in the tested cell lines. Furthermore, we affirm that chaperoning function is especially crucial for the stability of human telomeric and sub-telomeric regions during replication stress, thus reducing the catastrophic DSB occurrence at subtelomeres and telomeres. Our study, thus, reveals that HMGA2 protects the genome particularly the subtelomeric and telomeric stability by functioning as a global regulator of unconstrained supercoiled DNA. We henceforth propose that such a protective function of HMGA2, which initially evolved to protect ESC integrity, has been acquired by cancer cells to evade DNA-damaging chemotherapeutic drugs. Our findings, therefore, suggest that the interplay between HMGA2 and Topoisomerase might have far-reaching implications in personalized cancer therapy.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	i
ABSTRACT.....	iv
LIST OF FIGURES	xi
LIST OF TABLES.....	xvi
ACRONYMS.....	xvii
INTRODUCTION	1
1.1 Telomeres.....	1
1.1.1 An overview of the telomeres	1
1.1.2 The Subtelomeres.....	2
1.1.3 Telomere length maintenance by Telomerase	3
1.1.4 The Shelterin complex	4
1.1.5 Replication of telomeres	5
1.1.6 Telomeres and cancer	8
1.2. High Mobility Group (HMG) proteins	10
1.2.1 An overview of the HMG proteins	10
1.2.2 The High Mobility Group AT hook (HMGA) protein sub family.....	10
1.2.3 Phenotypes associated with HMGA proteins	12
1.2.4 HMGA and cancer	13
1.2.5 HMGA in DNA repair	14
1.2.6 Role of HMGA2 at the stalled replication forks.....	15
1.3. Stalled replication fork remodelling and the pathological chicken-foot structure	17
1.3.1 Role of Topoisomerases in relieving topological stress	18
1.3.2 HMGA2 and DNA topology.....	21
1.3.3 HMGA2 and Telomeres.....	22
1.4. The CRISPR-Cas system	23
1.5 OBJECTIVES OF THIS STUDY.....	25
2. MATERIALS AND METHODS.....	27
2.1 Endogenous tagging of TRF1 using CRISPR/Cas9	27

2.1.1 Design of the sgRNA	27
2.1.2 Design of the repair template	28
2.1.3 Co-transfection and selection of positive clones	29
2.1.4 Characterization of the selected clones	30
2.2 Cell culture	31
2.3 Genomic DNA Extraction	32
2.4 Agarose gel electrophoresis	33
2.5 Pulsed-Field Gel Electrophoresis (PFGE)	33
2.6 Terminal Restriction fragment (TRF) analysis	34
2.7 Southern blotting	34
2.8 PCR analysis	34
2.9 Sequencing of the PCR products	36
2.10 TA-cloning	36
2.11 Immunoblotting	36
2.12 Cytotoxicity assays:	38
2.13 Immunostaining	38
2.14 Optogenetic oxidative stress induction	39
2.15 Next Generation Sequencing Analysis	39
3. RESULTS	42
Chapter 1	42
3.1 CRISPR/Cas9 mediated tagging of TRF1 in HT1080 and hESCs	42
3.1.1 Selection of the sgRNA	42
3.1.2 Design of the repair template for knock-in	44
3.1.3 Co-transfection of the sgRNA/Cas9 vector and GFP-FLAG donor plasmid in HT1080 cells	45
3.1.4 Generation of clonal population from the targeted HT1080 cells	45
3.1.5 Imaging and PCR screening of the targeted HT1080 clones	46
3.1.6 Screening for potential bi-allelic gene targeting event in the tagged HT1080 clones	50
3.1.7 Southern blot analysis of the integrated HT1080 clones	51
3.1.8 Immunoblotting of the tagged TRF1 protein in HT1080 clones	54
3.1.9 Co-localization study with the derived HT1080 clones	55

3.1.10 Endogenous tagging of hESCs TRF1 with GFP-FLAG and mCherry-miniSOG using CRISPR/Cas9:	56
3.1.11 Imaging and PCR screening of the targeted hESC clones.....	57
3.1.12 Screening of potential bi-allelic gene targeting event in hESC clones.....	60
3.1.13 Southern blot analysis of the targeted hESC clones	61
3.1.14 Immunoblotting of the tagged TRF1 protein with anti-FLAG and anti-TRF1 antibody in targeted hESC clones.....	65
3.1.15 Subcloning of hESC-TRF1-mCherry-miniSOG clone hESC ^{C17}	67
3.1.16 Imaging and PCR screening of the hESC ^{C17} subclones	68
3.1.17 PCR to screen for potential mono/bi-allelic targeting in the hESC ^{C17} subclones.....	71
3.1.18 Immunoblotting to detect fusion TRF1 protein expression in hESC ^{C17} subclones.....	73
3.1.19 DNA damage induction by miniSOG	73
Chapter 2.....	77
3.2.1 Long-term knockdown of HMGA2 does not affect Telomere length	77
3.2.2 Role of HMGA2 at telomeres during HU induced replication stress	78
3.2.3 HMGA2 protects HU induced stalled replication forks from collapsing into DSBs at telomeres and subtelomeres.....	80
3.2.4 HMGA2 prevents collapse of HU induced stalled forks genome-wide in HT1080 C1 and C2 cells.....	82
3.2.5 HMGA2 prevents HU induced stalled forks against collapse genome-wide in NSCLC H1299 and A549 cells.....	83
3.2.6 Differential chemosensitivity of HMGA2 expressing cells against Topoisomerase I poison (SN38) induced DNA damage.....	86
3.2.7 HMGA2 protects H1299 and A549 cells against SN38 induced DNA damage	86
3.2.8 HMGA2 sensitizes HT1080 C1 and C2 cells against SN38 induced DNA damage	88
3.2.9 Differential role of HMGA2 in SN38 induced chemosensitivity impacts cell survival.....	89
3.2.9.1 HMGA2 prevents apoptosis in SN38 treated A549 cells	89
3.2.9.2 HMGA2 decreases cell viability in SN38 treated HT1080 C1 cells	90

3.2.10 Expression of HMGA2 does not alter the endogenous expression of Topoisomerases.....	91
3.2.11 Differential expression of HMGA2 determines the differential chemosensitivity to Topoisomerase I poison (SN38).....	94
Chapter 3.....	98
3.3.1 Subtelomeric regions are hypersensitive to replication stress and under the control of HMGA2.....	98
3.3.2 Ex vivo analysis to validate the hypersensitivity of subtelomeres and telomeres towards replication stress	106
4. DISCUSSION AND FUTURE PROSPECTS.....	109
4.1 Generation of TRF1 tagged stable cell lines using CRISPR/Cas9 genome editing tool	109
4.2 Effect of oxidative stress at telomeres (proof of concept for miniSOG tagged hESC-TRF1 stable cell line)	110
4.3 HMGA2 mediated fork protection of the stalled forks under replication stress	111
4.4 Mechanism of the protective role of HMGA2 at stalled replication forks	114
4.5 In silico and ex vivo analysis to determine the genomic locus of HMGA2-associated regions vulnerable to replication stress	119
5. CONCLUSION.....	122
6. REFERENCES	124

LIST OF FIGURES

- Figure 1: A schematic illustration of a mammalian chromosome showing the centromeres, telomeres and subtelomeres.
- Figure 2: A graphical illustration of the telomeres coated with the six-protein Shelterin complex.
- Figure 3: Figure showing the natural replication fork barriers encountered by a replication fork at telomeres.
- Figure 4: Consequences of telomere dysfunction in cancer and aging.
- Figure 5: Schematic representation of HMGA genes and the corresponding encoded protein.
- Figure 6: Model of HMGA2 mediated fork protection at stalled replication forks.
- Figure 7: Schematic diagram showing the process of formation of fork reversal.
- Figure 8: Diagram showing the mechanism of three-step relaxation of supercoils by Topoisomerase I.
- Figure 9: Working principle of CRISPR/Cas9 endonucleases.
- Figure 10: Graphical representation of steps involved in sgRNA selection.
- Figure 11: Donor plasmid of GFP-FLAG template.
- Figure 12: Donor plasmid of mCherry-miniSOG template.
- Figure 13: Graphical representation of the steps involved in CRISPR/Cas9 knock-in strategy performed in the study.
- Figure 14: Schematic representation of the different sgRNAs chosen in silico.
- Figure 15: SURVEYOR assay to measure the ability of each sgRNA to facilitate targeted cleavage.
- Figure 16: Schematic diagram of the GFP-FLAG and mCherry-miniSOG donor plasmid construct used for knock in at TRF1 locus.
- Figure 17: Confocal image analysis of the bulk cells expressing GFP-TRF1.
- Figure 18: PCR amplification of the Left Homology arm integration (LHA) in HT1080 clones.
- Figure 19: PCR amplification of the Right Homology arm (RHA) integration in HT1080 clones.
- Figure 20: Confocal microscopy image of HT1080 clones HT1080^{G#A2} and HT1080^{G#D11} showing GFP expression at TRF1 locus.

Figure 21: PCR amplification and sequence verification of the donor plasmid integration in HT1080 clones.

Figure 22: Schematic representation of the PCR amplification to screen for monoallelic or bi-allelic targeting events.

Figure 23: PCR amplification of the untargeted allele in HT1080 clones and the sequencing outcome.

Figure 24: Southern blot analysis of targeted HT1080 clones with internal GFP and puromycin probe.

Figure 25: Western Blot analysis to detect the endogenous as well as fusion TRF1 in targeted HT1080 clones probed with anti-TRF1 antibody.

Figure 26: Western Blot analysis for fusion TRF1 in targeted HT1080 clones probed with anti-FLAG antibody.

Figure 27: Expression of GFP-FLAG tagged TRF1 (green), immunostaining of the endogenous TRF2 (red) and the co-localization of TRF1 and TRF2 foci in HT1080 clones.

Figure 28: Schematic diagram of a) GFP-FLAG donor plasmid construct and b) mCherry-miniSOG donor plasmid construct used in this study.

Figure 29: Confocal microscopy images showing GFP signal for hESCs clones hESC^{G41}, hESC^{G46}. Corresponding DAPI stain images are shown.

Figure 30: Confocal microscopy images showing mCherry signal for clone hESC^{C17}. Corresponding DAPI stain image is shown.

Figure 31: PCR amplification of the Left Homology arm (LHA) integration in hESC clones.

Figure 32: PCR amplification of the Right Homology arm (RHA) integration in hESC clones.

Figure 33: PCR amplification to screen for potential bi-allelic targeting of the selected hESC clones.

Figure 34: Sequence alignment of the untargeted allele of clones hESC^{G41}, hESC^{G46}, hESC^{G55} and hESC^{C17} to the WT hESC.

Figure 35: Southern blot analysis of targeted hESC clones with an internal GFP probe.

Figure 36: Southern blot analysis of targeted hESC clones with an external probe.

Figure 37: A Western Blot analysis for fusion TRF1 in targeted hESC clones probed with anti-FLAG antibody.

Figure 38: A Western Blot analysis for fusion TRF1 in targeted hESC clones probed with anti-TRF1 antibody.

Figure 39: Schematic diagram of the mCherry-miniSOG donor plasmid construct used in this study.

Figure 40: Confocal images of hESC^{C17} subclones C17-A1, C17-A3, C17-A7 and C17-B1 expressing mCherry-miniSOG.

Figure 41: PCR amplification of the Left (LHA) and Right (RHA) Homology arms integration in hESC^{C17} subclones.

Figure 42: PCR amplification for sequence verification of the integration cassette in hESC^{C17} subclones.

Figure 43: PCR amplification to verify bi-allelic targeting in hESC^{C17} subclones.

Figure 44: Sequencing outcome of the untargeted allele PCR product in hESC^{C17} subclones.

Figure 45: A Western Blot analysis for TRF1 in targeted hESC clones probed with anti-TRF1 antibody.

Figure 46: Cell survival assay performed in hESC miniSOG expressing clone C17-A3 and hESC WT cells after exposure to blue light.

Figure 47: Cell survival assay performed in hESC miniSOG expressing clone C17-A3 and hESC WT cells of each row for different exposure times to blue light.

Figure 48: TRF analysis after long-term knockdown of HMGA2 using dox treated HT1080 C1 cells compared to dox treated HT1080 WT cells.

Figure 49: An immunoblot showing the down-regulation of HMGA2 in HT1080 C1 and C2 cells via doxycycline inducible shRNA.

Figure 50: TRF analysis of the HMGA2 expressing HT1080 C1 (Dox-) and HMGA2 knockdown (Dox+) cells treated with and without HU (HU+/ HU-).

Figure 51: Southern blot analysis with telomere probe of HU treated HT1080 C1 cells in presence or absence of HMGA2 to analyze the DSBs occurring at telomeres.

Figure 52: PFGE analysis followed by EtBr staining of HT1080 C1 and C2 cells to measure the DSBs after HU treatment.

Figure 53: Sequence alignment of the TA-cloned PCR products of H1299^{HMGA2-KO} cell line.

Figure 54: An immunoblot for HMGA2 in H1299^{Parental} and H1299^{HMGA2-KO} cell lines.

Figure 55: PFGE analysis and EtBr staining of parental H1299^{Parental} and H1299^{HMGA2-KO} cells to measure the DSBs after HU treatment.

Figure 56: An immunoblot of HMGA2 in A549, 1.3, 1.5 and 1.6 cell lines.

Figure 57: Ethidium bromide staining of PFGE analysis of DSBs of A549 and HMGA2 expressing cell lines treated with different doses of HU.

Figure 58: EtBr staining of PFGE analysis to analyze genomic fragments of H1299^{Parental} and H1299^{HMGA2-KO} cell lines after SN38 treatment in a dose dependent manner.

Figure 59: Ethidium bromide staining of PFGE analysis to analyze DSBs in HMGA2 nil (A549) and HMGA2 expressing (1.3, 1.5 and 1.6) cells treated with SN38.

Figure 60: Ethidium bromide staining of PFGE analysis of HT1080 C1 and C2 HMGA2 expressing (Dox-) and HMGA2 knockdown (Dox+) cells after SN38 treatment in a dose dependent manner.

Figure 61: Colorimetric Caspase 3/7 activity assay, performed on HMGA2 nil A549 and HMGA2 expressing 1.3, 1.5 and 1.6 cells when treated with SN38.

Figure 62: Cell survival assay, performed in HMGA2 expressing (Dox-) and HMGA2 knockdown (Dox+) HT1080 C1 cells treated with indicated doses of SN38.

Figure 63: Immunoblots of Topo1B in HT1080 C1/C2 and A549 cell line and HMGA2 expressing clones.

Figure 64: Immunoblots of Topo2A in HT1080 C1/C2 and A549 cell line and HMGA2 expressing clones.

Figure 65: Immunoblots of Topo2B in HT1080 C1/C2 and A549 cell line and HMGA2 expressing clones.

Figure 66: Comparative immunoblot of HMGA2 expression level across all the cell lines used in the study.

Figure 67: Comparative immunoblot of endogenous Topo 1B and HMGA1 expression level across all cell lines used in the study.

Figure 68: Ethidium bromide staining of the PFGE of SN38 treated H1299^{Parental} and H1299^{HMGA2-KO}.

Figure 69: Distribution of the enriched bins in SN38 induced DSB fragments of H1299^{Parental} (left panel) and H1299^{HMGA2-KO} (right panel) samples.

Figure 70: Coverage plot (chr17 shown as representative).

Figure 71: Enrichment distribution of the first/last 100kb of H1299^{Parental} (left panel) and H1299^{HMGA2-KO} (right panel) SN38 versus untreated samples.

Figure 72: Circos plot showing genome-wide distribution of differential enrichment and depletion in the SN38 induced DSB fraction of H1299^{Parental} and H1299^{HMGA2-KO} cell lines.

Figure 73: Locus of SN38 induced DSBs at the telomeres and subtelomeres. UCSC tracks (500bp bin) of SN38 treated H1299^{Parental}/H1299^{HMGA2-KO} and untreated H1299^{Parental}/H1299^{HMGA2-KO} cells.

Figure 74: Southern blot with telomere probe of H1299^{Parental} and H1299^{HMGA2-KO} cells with indicated doses of SN38 showing the DSB occurrence at subtelomeric and telomeric regions.

Figure 75: Southern blot with telomere probe in HT1080 C2 HMGA2 expressing (Dox-) and HMGA2 knock down (Dox+) cells with indicated doses of SN38 showing the DSB occurrence at subtelomeric and telomeric regions.

Figure 76: Schematic diagram representing the DSB fragments occurring at telomeric and subtelomeric regions after induced replication stress.

Figure 77: Proposed mechanism showing the role of HMGA2 in HU induced replication stress.

Figure 78: Proposed mechanism showing the role of HMGA2 in SN38 induced replication stress.

LIST OF TABLES

Table 1: List of primers (5'-3') used in this study

Table 2: List of antibodies used in the study

Table 3: A table showing the intensity of each band and the percentage of indels created with different sgRNAs

ACRONYMS

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
DSBs	double strand breaks
nt	nucleotide
kb	kilobase
SD	segmental duplication
NHEJ	Non-Homologous End Joining
TERRA	Telomeric repeat-containing RNA
DDR	DNA damage response
TRF1	Telomeric repeat binding factor 1
TRF2	Telomeric repeat binding factor 2
OCT3/4	octamer-binding transcription factor 3/4
ATM	ATM serine/threonine kinase
ssDNA	single-stranded DNA
dsDNA	double-stranded DNA
RPA	Replication Protein A
ATR	ATR serine/threonine kinase
UPF1	up-frameshift suppressor
D-loop	Displacement Loop
T-loop	Telomere Loop
NM	nuclear matrix
CST complex	CTC1, STN1, and TEN1 complex
WRN helicase	Werner syndrome RecQ like helicase
FEN1	Flap endonuclease 1
RECQ	RecQ Like Helicase
RTEL1	regulator of telomere elongation helicase 1
MTS	multi-telomere signals
FISH	fluorescent in-situ hybridization
BLM helicase	Bloom syndrome RecQ like helicase
ALT	Alternate Lengthening of Telomeres
KO	Knockout

KDa Kilo Dalton
WT Wild Type
SAHF Senescence-associated heterochromatin foci
NSCLC non small cell lung cancer
BER base excision repair
CHK1 Checkpoint kinase 1
BRCA1 breast cancer 1
hESC Human Embryonic Stem Cells
dNTP Deoxyribonucleotide triphosphate
PCNA Proliferating cell nuclear antigen
HR Homologous Recombination
AFM Atomic force microscopy
Co-IP Co-immunoprecipitation
TIFs Telomere dysfunction- Induced Foci
CRISPR Clustered regularly interspaced short palindromic repeats
Cas CRISPR associated
sgRNA short guide RNA
PAM Protospacer Adjacent Motif
HDR Homology Directed Repair
GFP Green Fluorescent Protein
miniSOG mini Singlet Oxygen Generator
TRF1 gene Telomeric Repeat Binding Factor 1
PCR Polymerase chain reaction
TBE Tris/Borate/EDTA
EDTA Ethylenediaminetetraacetic acid
gDNA Genomic DNA
TE Tris-EDTA
EtBr Ethidium Bromide
PBS Phosphate-buffered saline
SDS Sodium dodecyl sulfate
SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis
PVDF Polyvinylidene fluoride
HRP horseradish peroxidase
ECL enhanced chemiluminescent substrate

PFGE Pulsed-Field Gel Electrophoresis
DIG Digoxigenin
GRCh38 Genome Reference Consortium Human Build 38
UCSC University of California, Santa Cruz
FACS Fluorescence-activated cell sorting
ROS reactive oxygen species
KD/kd knockdown
MTL mean telomere length
TRF Terminal Restriction Fragment
Dox Doxycycline
HU Hydroxyurea
scDNA supercoiled DNA
Top1cc Topoisomerase-1-DNA cleavage complexes
PDB protein-linked DNA break
PDX Patient derived xenograft

INTRODUCTION

1.1 Telomeres

1.1.1 An overview of the telomeres

Telomeres are specialized nucleo-protein complexes at the ends of the linear chromosomes that are crucial for the genome stability. They protect the integrity of chromosomes by preventing the loss of vital coding sequences of genes at each cell cycle. Telomeres differentiate the chromosome termini from a broken DNA thus averting illegitimate fusions and recombination. Without a mechanism to maintain telomere length, telomeres shorten continuously in successive population doublings. After a typical number of cell divisions known as the Hayflick limit (Hayflick and Moorhead 1961), the cells undergo senescence which is generally associated with dysfunctional telomeres (d'Adda di Fagagna, Reaper et al. 2003). Telomere de-protection is catastrophic for cellular life span as it results in activation of DNA damage response leading to apoptosis (Ancelin, Brun et al. 1998, Chin, Artandi et al. 1999, Takai, Smogorzewska et al. 2003). In fact the length of telomeres determines the number of replication cycles of a cell and the timing of senescence (Kaul, Cesare et al. 2011).

Figure 1: A schematic illustration of a mammalian chromosome showing the centromeres (indicated in brown), telomeres (indicated in yellow) and subtelomeres (indicated in dark blue). The duplex human telomeric repeat sequence TTAGGG ranging 3-20 kb terminating in a 50-150 nucleotide 3' single stranded overhang are shown within the box.

Telomeres are highly conserved stretches of G rich tandem repeat sequences with minor variation between species (Blackburn and Gall 1978, Meyne, Ratliff et al. 1989, Pryde, Gorham et al. 1997). The lengths of telomeres vary from one species to another and from chromosome to chromosome. The shortest telomeres (36nt) are found in ciliates while the longest is in mice (ranging from 100 to 150kb). In humans, telomeres range between 3 to 20kb with arrays of TTAGGG (Figure 1) (Meyne, Ratliff et al. 1989, Kipling and Cooke 1990). A common trait of telomeres across species is the single-stranded overhang at the G rich 3' end of the chromosome (Figure 1). The length of the overhang also varies between species, which can be as short as 14nt in ciliates and 50-150nt in mammalian cells (Chai, Du et al. 2006). Like the rest of the chromosome telomeres are organized into chromatin. Electron microscopy images show that telomeres are organized in 30nm fiber structure (Nikitina and Woodcock 2004).

1.1.2 The Subtelomeres

Upstream of the telomeres towards the center of the chromosome, are the regions referred as the subtelomeres (Figure 1). They comprise ~5% of the human genome and is estimated to be >200kb, but vary in length depending on cell type and chromosome (Riethman, Ambrosini et al. 2004). Like telomeres, the subtelomeric region is also composed of unique sequence composition containing long arrays of repetitive sequences including that of telomeric repeat tracts (Riethman, Ambrosini et al. 2004). Characteristic feature of the subtelomers are the segmental duplication (SD) mosaics known as the 'Subtelomeric Repeats' which comprise 80% of the most distal 100kb in human DNA (Riethman 2008). Non-homologous end joining (NHEJ)-mediated translocations are reported to be responsible for such high rates of SD and has been speculated to have contributed to primate and human evolution (Linardopoulou, Williams et al. 2005, Riethman 2008, Tashiro, Nishihara et al. 2017). It has been established that the subtelomeres host the promoters for telomere transcription into TERRA (telomeric repeat-containing RNA) (Porro, Feuerhahn et al. 2014). Subtelomeric regions are reported to be involved in key biological activities such as, cell cycle regulation, heterochromatin formation at the chrosomal ends and transcriptional regulation of the subtelomeric genes (Riethman, Ambrosini et al. 2004,

Deng, Wang et al. 2012). Microdeletions in subtelomeric region has been implicated in diseases (De Vries, Winter et al. 2003, Stadler, Rahimov et al. 2013). Interestingly, a recent study reported p53-binding sites located at the subtelomeres in close proximity to telomeres suggesting the role of subtelomeres in suppressing the accumulation of DNA damage at the telomeres (Tutton, Azzam et al. 2016).

1.1.3 Telomere length maintenance by Telomerase

The ability of the DNA polymerase to synthesize DNA only in 5' to 3' direction leads to the 5' terminal gap called as the end-replication problem, due to which telomeres gradually shorten after every round of replication (Olovnikov 1973). In order to circumvent the loss of terminal DNA, a ribonucleoprotein called Telomerase functions as a reverse transcriptase. The catalytic subunit of telomerase is called the telomerase reverse transcriptase (TERT) and the RNA template is called the telomerase RNA (TR). Although telomerase has seven motifs similar to that of a conventional reverse transcriptase (Lingner, Hughes et al. 1997), its repeat addition processivity through rebinding of the RNA template on the same DNA primer makes it unique. Telomerase is highly expressed in stem cells and other germline cells but its expression is reduced in somatic cells to the extent that it is insufficient to compensate for telomere loss during each cell cycle ultimately leading to senescence (Bodnar, Ouellette et al. 1998). On the other hand, Telomerase is re-expressed in tumor cells where it is directly associated with immortality of cells. Up to 80% of cancer cells show telomerase up-regulation after transformation (Harley 2008).

1.1.4 The Shelterin complex

Figure 2: A graphical illustration of the telomeres coated with the six-protein Shelterin complex. The shelterin proteins interact with the telomeric DNA directly (TRF1, TRF2 and POT1) or indirectly (TIN2, TPP1 and RAP1). The double stranded DNA binding proteins TRF1 and TRF2 and single stranded DNA binding protein POT1 are depicted.

Telomeres are nucleo-protein complexes that are bound by the Shelterin (Figure 2). Together, shelterin and telomeres ‘cap’ the chromosome ends protecting it from degradation by exonucleases, DNA-damage response (DDR) and DNA repair (O’Sullivan and Karlseder 2010). Shelterin is a protein complex with six subunits: Telomeric repeat binding factor 1 and 2 (TRF1 and TRF2), TRF1 interacting protein (TIN2), protection of telomeres protein 1 (POT1), TIN2 and POT1 interacting protein (TPP1) and repressor/activator protein1 (RAP1) (Figure 2) (de Lange 2005). Out of these, TRF1 and TRF2 are double-stranded DNA binding homodimeric proteins while POT1 is bound to the 3’ single-strand overhang (5’-TTAGGGTTA-3’). TRF1 and TRF2 are tethered by TIN2 to TPP1 hence often referred as the ‘linchpin’ (Ye, Donigian et al. 2004). TPP1 links the double stranded region to the single-stranded overhang by binding to POT1 and TIN2. Rap1 forms a complex with TRF2 (de Lange 2005).

Shelterin proteins are the hubs for a network of several other proteins that interact with telomeres. They have pivotal roles in protecting the telomeres as well as orchestrating faithful telomeric replication, which is crucial for genome integrity.

For instance, studies have shown that TRF1 suppresses telomere fragility and the activation of DDR and telomeres (Martinez, Thanasoula et al. 2009, Sfeir,

Kosiyatrakul et al. 2009). TRF1 has also been reported to take part in the semi conservative replication of telomeres (Sfeir, Kosiyatrakul et al. 2009). TRF1 is highly expressed in pluripotent cells and the expression of pluripotent markers OCT3/4, up-regulate TRF1, suggesting its role in telomere length maintenance (Schneider, Garrobo et al. 2013).

TRF2 represses the activation of DDR pathway ATM (Denchi and de Lange 2007). TRF2 is known to interact with Apollo, which is a 5'-3' exonuclease in the ATM pathway, disruption of which leads to increased replication stress (Lenain, Bauwens et al. 2006, van Overbeek and de Lange 2006). Through TRF2 interaction, Apollo and Topoisomerase II alpha function together to relieve the topological stress built up during telomere replication (Ye, Lenain et al. 2010).

Shelterin protein RAP1 suppresses NHEJ (Non-Homologous End Joining) at telomeres thus protecting the telomeres from illegitimate fusions (Sfeir, Kabir et al. 2010). POT1 displaces RPA from 3' overhangs and ssDNA formed during telomere replication (Gong and de Lange 2010). This is essential because excessive RPA is a sign of stalled fork and signals the activation of DNA damage response ATR (Denchi and de Lange 2007).

Through its interaction with TPP1, the helicase UPF1 (up-frameshift suppressor) displaces TERRA which can otherwise form R-loops causing replication stress as they are prone to collapsing into DSBs (Azzalin, Reichenbach et al. 2007, Azzalin 2012, Flynn, Cox et al. 2015).

1.1.5 Replication of telomeres

A bulk of telomeres are replicated by the conventional semi-conservative replication machinery (Pfeiffer and Lingner 2013). Although some studies suggest that replication initiates within the telomeric region and have traced firing origins within telomeric sequences, majority of studies have documented that telomeric replication forks move in an unidirectional manner initiated from the last subtelomeric origin

proximal to telomeres (Miller, Rog et al. 2006, Gilson and Geli 2007, Arnoult, Schluth-Bolard et al. 2010, Drosopoulos, Kosiyatrakul et al. 2012).

Replication timing of telomeres is known to differ among species. Human telomeres are known to replicate throughout the S phase whereas the subtelomeres replicate during the late S phase (Ten Hagen, Gilbert et al. 1990, Ofir, Wong et al. 1999, Wright, Tesmer et al. 1999, Hultdin, Gronlund et al. 2001). Replication timing of the human telomeres is chromosome arm-specific and depends on subtelomeric sequences where the origins for telomere replication are located (Arnoult, Schluth-Bolard et al. 2010).

Faithful replication of DNA is a key contributor towards genome stability. During each round of replication, a fraction of replication forks stall due to various exogenous factors such as DNA damaging agents or endogenous factors such as replication fork barriers (Mirkin and Mirkin 2007). If left unresolved, the stalled forks could result in irreversible fork collapse consequently leading to increased DNA double strand breaks (DSBs). Defects in the replication pathways can lead to mutagenesis or chromosomal rearrangements which are major causes of genomic instability (Aguilera and Gomez-Gonzalez 2008). Due to its unusual characteristics, telomeres pose numerous challenges for replication fork progression.

Figure 3: Figure showing the natural replication fork barriers encountered by a replication fork at telomeres such as the RNA: DNA hybrid termed as the R-loop, higher order heterochromatin formation, inherent torsional stress at the telomeres during replication, invasion of the single-stranded overhang into the double strand resulting in the telomere T-loop, the tucking in of the overhang at the invasion site forming the displacement D-loop and the higher order secondary structure called G-quadruplex formed by the G rich single stranded telomeric DNA.

Firstly the formation of loops at the ends of linear chromosomes creates an obstacle for the progression of replication fork. Telomeric DNA folds back several kilobases upstream into the double stranded DNA to form the T-loop where the 3' overhang tucks between the dsDNA, forming the D-loop (Figure 3) (Greider 1999, de Lange 2005). For the replisome to progress, the T-loop and D-loop have to be resolved. Furthermore, the nature of the t-loop where it is bound by protein complexes organized around a four-way junction like D-loop limits the free rotation of the loop, hence contributing as a topological barrier (Figure 3) (Gilson and Geli 2007).

Secondly, during replication when the telomeric DNA is separated by helicases, the single stranded G rich strand is prone to the formation of G-quadruplex (G4 DNA) through the hydrogen bond created by Hoogsteen base pairing of the G in TTAGGG (Figure 3). The G4 DNA formed on the lagging strand template could lead to replication fork arrest eventually leading fork breakage and telomere loss (Mirkin and Mirkin 2007, Oganessian and Bryan 2007).

Thirdly, the heterochromatic nature of telomeres due to interactions between the shelterin proteins and non-shelterin proteins as well the telomeric DNA generates topological constraints thus causing a barrier for the replisome (Figure 3) (Bandaria, Qin et al. 2016, Maestroni, Matmati et al. 2017).

Telomeres are transcribed into TERRA which initiates from the subtelomeric region and proceeds through the telomeric tracts (Azzalin, Reichenbach et al. 2007, Schoeftner and Blasco 2008). Under certain circumstances, when the RNA molecule anneals to its DNA template, it could lead to the generation of RNA:DNA hybrids. The strand displacement by the RNA:DNA hybrids typically forms a structure known as an R-loop (Costantino and Koshland 2015). The TERRA R-loops formed at the telomeres are considered to be natural fork barriers contributing to replication stress (Figure 3).

Moreover, recent studies have reported that human telomeres are attached to the nuclear matrix (NM) aided by shelterin and lamins (Luders, van Steensel et al. 1996, Burla, La Torre et al. 2016). As a result, when the replication fork approaches the telomeres, they are not free to rotate to dissipate the torsional stress, resulting in

increased local topological strain. Several studies have shown that telomeres are exceptionally vulnerable to topological stress (Germe, Miller et al. 2009, Ye, Lenain et al. 2010, Chen, Zhu et al. 2015, Hsieh, Tsai et al. 2015, Benarroch-Popivker, Pisano et al. 2016).

For these reasons, telomeres are considered as fragile sites and difficult-to-replicate regions of the genome. Shelterin proteins especially TRF1 and TRF2 are known to play pivotal roles in aiding telomere replication (Lenain, Bauwens et al. 2006, van Overbeek and de Lange 2006, Sfeir, Kosiyatrakul et al. 2009). In addition to the standard replication machinery, numerous non-shelterin proteins also interact with telomeres through shelterin to ensure efficient replication. These proteins include, the human CST complex (Stewart, Chaiken et al. 2012), BLM helicase (Drosopoulos, Kosiyatrakul et al. 2015) WRN helicase (Edwards, Orren et al. 2014), Apollo exonuclease (Lenain, Bauwens et al. 2006, van Overbeek and de Lange 2006), Topoisomerase II alpha (Ye, Lenain et al. 2010), structure specific endonuclease FEN1 (Saharia, Teasley et al. 2010), DNA repair protein called RECQ1 (Sami, Lu et al. 2015) and RTEL1 helicase (Ding, Schertzer et al. 2004). Depletion of any such telomere maintenance proteins, leads to telomere fragility observed as multi-telomere signals (MTS) on metaphase chromosomes by fluorescent *in situ* hybridization (FISH).

1.1.6 Telomeres and cancer

Among the several telomere-associated diseases, cancer has gained much attention due to its complexity and prevalence. Several telomere-aberration events have been reported to take place during the development of cancer (Figure 4). Unlike normal cells that arrest after reaching critically short telomere length, cancer cells can bypass the checkpoints and maintain telomere length by activating telomerase (Wright and Shay 1992). Furthermore, senescence is triggered in normal cells by the activation of p53 when five telomeres become dysfunctional. By attaining multiple mutations that lead to loss of function of tumor suppressors, cancer cells evade senescence (Figure 4) (Kaul, Cesare et al. 2011, Cesare and Karlseder 2012). ~90% cancer cells maintain telomere length by reactivating telomerase. However, certain tumor cells employ a different mechanism called ALT (Alternate Lengthening of Telomeres) to elongate

the telomeres. ALT lengthens telomeres by homology dependent recombination and RAD51 is reported to be involved (Pickett and Reddel 2015, Maciejowski and de Lange 2017).

Mutations in Shelterin proteins are implicated in tumorigenesis. TRF1/p53 double KO mice immediately develop invasive squamous cell carcinoma suggesting the role of TRF1 as a tumor suppressor (Martinez, Thanasoula et al. 2009). TRF2 expression is found to be elevated in human skin carcinomas (Munoz, Blanco et al. 2005) and upregulation of TRF2 resulted in the onset of ALT pathway (Blanco, Munoz et al. 2007). RAP1 is known to interact with the telomeric repeats in the subtelomeric region and control subtelomeric gene expression (silencing). RAP1 mutations have been reported in cancer suggesting a role in epigenetic modifications after onset of cancer (Martinez, Thanasoula et al. 2010).

Figure 4: Consequences of telomere dysfunction in cancer and aging (Martinez and Blasco 2010).

Telomere dysfunction owing to critically short telomeres or the uncapping of shelterin complex stimulates DNA damage response pathways by the activation of kinases such as DNA-PK, ATM or ATR. In the first possible outcome, the damaged DNA is repaired by NHEJ or HR, inducing terminal fusions or illegitimate recombination eventually leading to genomic instability and cancer initiation. In the second outcome, activation of tumor suppressors such p53 or p21 leads to programmed cell apoptosis or senescence leading to loss of function of tissues and organs and aging.

1.2. High Mobility Group (HMG) proteins

1.2.1 An overview of the HMG proteins

The High mobility group (HMG) proteins are non-histone nuclear DNA binding proteins. They were named owing to their small size (17 KDa) resulting in high electrophoretic mobility when separated on polyacrylamide gels and high content of acidic and basic amino acids. The HMG family comprises of HMGA, HMGB and HMGN proteins, which are non-homologous and differ in their DNA binding modes (Bianchi and Agresti 2005). The HMGA subfamily (previously known as HMGI) has four proteins the HMGA1a, HMGA1b and HMGA1c resulting from alternative splicing of the same HMGA1 gene and HMGA2 that is transcribed from another gene (Reeves and Beckerbauer 2001). HMGA1 and HMGA2 genes are located on chromosome 6p21.31 and 12q14.3 respectively (Figure 5) (Fusco and Fedele 2007).

1.2.2 The High Mobility Group AT hook (HMGA) protein sub family

The characteristic feature of the HMGA family is the presence of three short palindromic motifs called the 'AT-hooks' (Pro-Arg-Lys-Arg-Pro) through which they bind to the AT rich DNA sequences of the minor groove. While the basic N-terminus region is known for DNA interaction, the C-terminus acidic region (except in HMGA1c) is speculated to mediate protein-protein interactions (Figure 5) (Sgarra, Maurizio et al. 2009). The main difference among the HMGA proteins is the sequence variation in the N-terminus region and between the AT-hooks of the protein within the family (Fedele, Battista et al. 2001). HMGA2 and HMGA1 proteins have an overall sequence homology of 55% (Reeves 2000). These proteins are known to modulate transcription during development by binding and modulating the chromatin architecture. Hence HMGA proteins are regarded as the 'architectural' transcription factors. They are also known mediators for assembling multi-protein higher-order complexes to enable synergistic effects of transcription factors (Du, Thanos et al. 1993).

Figure 5: Schematic representation of HMGA genes and the corresponding encoded protein (Fusco and Fedele 2007). HMGA1 (HMGA1a, HMGA1b) (A) and HMGA2 (B) proteins are well conserved and have similar structure. The exons encoding corresponding proteins are indicated in blue. Each protein has three basic DNA-binding domains called AT-hooks indicated in green and a acidic C-terminal tail indicated in red.

HMGA proteins are highly expressed in embryonic stem cells and during early developmental stages (Chiappetta, Avantaggiato et al. 1996). They are mostly undetectable in somatic cells but their expression is restored in most benign and malignant tumors. Low expression of HMGA1 has been found in some somatic cells whereas HMGA2 expression has not been detected so far (Rogalla, Drechsler et al. 1996, Chase, Haga et al. 1999). Interestingly, HMGA2 is highly expressed in induced pluripotent stem cells (iPSc) (Morshedi, Ren et al. 2013).

While HMGA proteins are intrinsically unstructured in solution, upon binding to DNA substrate, the AT-hooks gradually transition to a crescent shaped structure (Huth, Bewley et al. 1997). Although they are known to bind to AT rich sequences, the main determinant of binding is the structure of the DNA substrate rather than the sequence. HMGA proteins are known to have affinity for four-way Holliday junction, synthetic chicken foot like structures, bent or supercoiled DNA and unpaired regions of the AT-rich DNA (Nissen and Reeves 1995, Hill and Reeves 1997, Reeves 2001, Zhao, Peter et al. 2017). Depending on the sequence, length or topology of the substrate DNA, HMGA proteins can bend, straighten, loop or unwind the DNA and furthermore induce supercoils in relaxed plasmid DNA *in vitro* (Falvo, Thanos et al.

1995, Nissen and Reeves 1995, Lehnert, Reat et al. 1998, Chase, Haga et al. 1999, Bagga, Michalowski et al. 2000, Zhao, Peter et al. 2017).

1.2.3 Phenotypes associated with HMGA proteins

Direct consequence of HMGA abnormality during development was shown in mice and chicken before it was reported in human. Mice with functional mutations in HMGA2 gene exhibit a 'pygmy' phenotype characterized by severe deficiency in fat cells, short head and reduced birth weight as well as dwarfism. They have 60% reduced body weight in the case of null-mice and 25% reduced body weight in the case of heterozygous KO compared to the WT littermates (Zhou, Benson et al. 1995). Also, male HMGA2 homozygous knockout mice were sterile (Chieffi, Battista et al. 2002). The dwarfism phenotype was also reported in HMGA2 knockout chicken (Ruyter-Spira, de Groof et al. 1998, Ruyter-Spira, Herbergs et al. 1998).

On the other hand, knockout of HMGA1 led to impaired adipocyte differentiation, cardiac hypertrophy, type II diabetes and formation of malignancies suggesting its role in the negative regulation of adipocytic cell growth. These mice have 27% reduced body weight compared to the wild type littermates (Foti, Chiefari et al. 2005, Fedele, Fidanza et al. 2006). Taken together these data suggest that HMGA1 and HMGA2 play distinct roles and the balance of them is essential for appropriate cell growth and proliferation.

The knockout of both HMGA1 and HMGA2 resulted in much smaller mice called 'super-pygmy' due to the severe dwarfism phenotype (Federico, Forzati et al. 2014). These mice were extremely small having 75% reduced body weight compared to the WT. While most of them die in utero, the surviving littermates die within the first 12 months, indicating the role of HMGA proteins in early stage vitality (Federico, Forzati et al. 2014).

The first case of HMGA2 haploinsufficiency was reported in a boy born with a micro deletion in chromosome 12q14.4. The observed phenotypes were pre- and post-natal

growth delay, short stature (dwarfism) with triangular face, mental retardation, reduced adipose tissue and muscular hypotrophy (Mari, Hermanns et al. 2009).

The over expression of HMGA2 as a result of *de novo* pericentric inversion on chr12 with breakpoints at p11.22 and q14.3 leading in a truncated HMGA2 gene was reported in a 8 year old boy. The phenotype included somatic over growth (gigantism), advanced bone and dental ages, arthritis, cerebellar tumor and multiple subcutaneous lipomas, increased fat tissues and facial dysmorphism (Ligon, Moore et al. 2005).

1.2.4 HMGA and cancer

The upregulation of HMGA proteins have often been observed in cancer classifying them as a ‘tumor marker’. The property of HMGA to set up a transcriptional network that maintains a stem-like state during development is observed in cancer. This upregulation of HMGA1 and HMGA2 is observed in both benign and malignant tumors and the expression correlates with cancer metastasis. The blockage of HMGA upregulation suppresses the tumor phenotype (Trapasso, Sarti et al. 2004). The aberrant restoration of HMGA protein expression after onset of cancer is associated with the relocations of the HMGA genes and its truncated version is characterized in tumor cells. Most often, benign tumors of mesenchymal origin have truncated HMGA2 gene resulting in the loss of its 3’ UTR, which possess the Let-7 binding site that represses HMGA2 (Kazmierczak, Wanschura et al. 1995, Kools and Van de Ven 1996). Overexpression of HMGA1 and HMGA2 has been found in pituitary adenomas as well. The overexpression of the two proteins in mice resulted in the transformation of benign tumor to malignant depicting a causal role of HMGA proteins in cancer progression (Fedele, Battista et al. 2002, Fedele, Pentimalli et al. 2005).

During senescence, HMGA proteins were shown to promote the formation of Senescence-associated heterochromatin foci (SAHF) suggesting a possible tumor-suppressor role. The same study also indicated that HMGA proteins are essential structural components of the senescence machinery (Narita, Narita et al. 2006). However, the majority of studies support its oncogenic role. Over expression of

HMGA1 and HMGA2 is a common feature in carcinogenesis including colon cancer (Abe, Watanabe et al. 1999, Wang, Liu et al. 2011), breast cancer (Chiappetta, Botti et al. 2004, Jones, Mitter et al. 2008), pancreatic cancer (Piscuoglio, Zlobec et al. 2012), ovarian cancer (Masciullo, Baldassarre et al. 2003), non small cell lung cancer (NSCLC) in both adenocarcinoma and squamous cell histotypes (Kettunen, Anttila et al. 2004, Rice, Lai et al. 2013, Zhang, Wang et al. 2015), esophageal cancer (Chen, Lechago et al. 2004) and testicular cancer (Franco, Esposito et al. 2008).

The common feature in these cancers apart from the over expression of HMGA proteins is that the expression of HMGA leads to poor prognosis, response to treatment and correlate with the metastatic potential leading to reduced survival (Fusco and Fedele 2007).

1.2.5 HMGA in DNA repair

Numerous studies have proposed a role of HMGA proteins in DNA repair. HMGA proteins are known to regulate the transcription of DNA repair proteins also directly participate in the DNA repair pathways (Adair, Kwon et al. 2005). In this regard, the increased expression of HMGA proteins during embryogenesis seems crucial for proper fetal development whereas in the case of cancer, it can cause an imbalance in DNA repair pathways thus promoting carcinogenesis. Consistent with this notion, both HMGA1 and HMGA2 are known to be involved in the ATM pathway, where their expression increases the expression level of ATM and vice versa (Trapasso, Sarti et al. 2004).

A previous study from our laboratory demonstrated that HMGA1 and HMGA2 proteins have an intrinsic dRP/AP site cleavage activity that affects the BER associated repair pathway and promotes cellular resistance against chemotherapy (Summer, Li et al. 2009). Several studies have shown that HMGA2 promotes cell survival by inhibiting apoptosis (Liu, Liang et al. 2015, Cai, Shen et al. 2016, Shi, Wu et al. 2016). This is shown to occur through its interaction with ATR-CHK1 complex in the cancer cells (Natarajan, Hombach-Klonisch et al. 2013). Studies have shown direct binding of HMGA1 to the promoter of tumor suppressor BRCA1 leading to the

downregulation of BRCA1 expression. BRCA1 is involved in homologous recombination DNA repair (Baldassarre, Battista et al. 2003).

1.2.6 Role of HMGA2 at the stalled replication forks

At each cell cycle, a fraction of replication forks stall due to various reasons like repeat sequences, secondary structures or damaged DNA templates (Mirkin and Mirkin 2007). The stalled forks serve as a signal for DNA repair pathways to facilitate fork repair and restart. In the case of prolonged stalling, forks become vulnerable to endonucleolytic attack and irreversibly collapse into DSBs (Branzei and Foiani 2010). Prolonged stalling of forks also triggers apoptotic signals, which are compromised in conditions such as carcinogenesis.

The first evidence for the role of HMGA2 directly interacting with an ongoing replication fork was reported in our laboratory (Yu, Lim et al. 2014). According to the study, HMGA2 forms a scaffold at the stalled forks and protects it against endonucleolytic attack, which would otherwise convert the forks into DSBs (Figure 6). It was shown that during induced replication stress, HMGA2 levels are enriched in the cells (Yu, Lim et al. 2014). Moreover, cells expressing HMGA2 exhibit resistance to induced replication stress whereas cells that do not express HMGA2 exhibit increased sensitivity. This effect was shown in *E. coli*, mouse ESC, human ESC, yeast cells and HT1080 human fibrosarcoma cell line (Yu, Lim et al. 2014).

The idea for the chaperoning effect of HMGA2 at stalled forks emerged because of the hypersensitivity of the HMGA2 knockdown cells to hydroxyurea (HU) treatment (Summer, Li et al. 2009). HU inhibits the ribonucleoside diphosphate reductase. As a consequence, once the dNTP pool is depleted, replication forks initially slow down and eventually stall, arresting the cells in S phase (Krakoff, Brown et al. 1968). The increased sensitivity of HMGA2 knockdown cells to HU indicated its possible role in the stabilization of stalled replication forks. The results demonstrated that HMGA2 co-localizes with two main replication factors RPA and PCNA with or without HU induced stress. This means that HMGA2 is present in close proximity to the ongoing as well as stalled forks. Furthermore, induced replication stress in the absence of

HMGA2 produced more γ H2AX foci, which indicates the activation of DNA damage signals pertaining to DSBs (Yu, Lim et al. 2014).

Figure 6: Model of HMGA2 mediated fork protection at stalled replication forks (Yu, Lim et al. 2014).

Stalled forks are substrates for endonucleolytic attack that results in irreversible fork collapse into DSBs. Accumulation of such DSBs leads to chromosomal instability and cell death. In cells that express HMGA2, the protein acts as a replication fork chaperone protecting the stalled forks from endonucleolytic attacks. This increases the likelihood of fork repair and restart eventually contributing to genome stability.

Experiments with HMGA2 and reconstituted Y-shaped Holliday Junction-like structures and the regressed fork structure often called ‘chicken foot’ showed that HMGA2 has a higher intrinsic affinity for these structures than to AT-rich duplex DNA. It was also shown that fork protection requires all three AT-hooks to be active. This is consistent with the previous finding that HMGA1a has higher affinity for such structures and requires all three active hooks for the stable binding to DNA (Hill and Reeves 1997).

Notably, an experiment in yeast revealed a possible mechanism for this fork protection effect. Recombinant HMGA2 expressed in yeast could partially rescue the stalled replication forks in *mec1* (ATR ortholog in yeast pivotal for fork stabilizing activity) and *sml1* (regulates dNTP production in yeast) deficient cells. Two-dimensional gel electrophoresis followed by Southern blot analysis to track the early-firing replicon ARS305 after HU challenge, revealed that HMGA2 restricts the formation of X-spike like structure, which signifies the four-way regressed fork (chicken-foot). Fork regression is a pathological structure that is more prone to collapse into DSBs (Budzowska and Kanaar 2009).

Overall it is likely that the fork chaperone function, which primarily evolved as a protective mechanism for ES cells, is hijacked by HMGA2 expressing tumor cells to evade chemotherapeutic drugs and DNA damaging agents (Yu, Lim et al. 2014).

1.3. Stalled replication fork remodelling and the pathological chicken-foot structure

A major challenge for faithful DNA replication is the frequent breakdown of replication forks by various endogenous and external factors causing genome instability and is also known to be associated with cancer formation (Aguilera and Gomez-Gonzalez 2008). Factors such as unrepaired DNA lesions, misincorporation of dNTPs, unusual DNA structures, common fragile sites, the conflict between replication and transcription machineries, nicks and gaps and topological stress also occur frequently, causing replication fork stalling (Zeman and Cimprich 2014).

To counteract these barriers, multiple pathways have evolved to protect the stalled forks and ensure restart to facilitate error-free replication. One of the first protective pathways that was proposed in 1976 suggested the remodelling of stalled forks to regressed fork like structure (Figure 7) (Higgins, Kato et al. 1976). This occurs when the nascent strands come together and anneal at the junction of the unreplicated parental strands reversing back in the direction opposite to the replisome resembling a ‘chicken-foot’ like structure. Although, this process has the potential to rescue the stalled fork by providing new repair template for HR, direct restart of forks bypassing

the lesion could also occur (Neelsen and Lopes 2015). Hence, the efficiency of the post-replication repair mechanism determines if the fork reversal was advantageous or detrimental. Studies have also shown that in the event of encountering such a lesion, the replisome either disassembles or moves away from the site (Cotta-Ramusino, Fachinetti et al. 2005) which could expose newly synthesized DNA to structure-specific endonucleases or exonucleases. This exposure could result in an irreversible fork collapse into DSBs, eventually leading to cell death. (Figure 7)

Figure 7: Schematic diagram showing the process of formation of fork reversal. Unwinding of duplex strands during semi-conservative replication leads to positive supercoil formation ahead of the fork which are resolved by the topoisomerases. In case of fork stalling, the supercoil accumulates resulting in fork remodelling where the nascent strands come together and anneal at the junction of the unreplicated parental strands reversing back in the direction opposite to the replisome resembling a ‘chicken-foot’ like structure.

1.3.1 Role of Topoisomerases in relieving topological stress

Within the cells, the topological state of the DNA is highly dynamic and rapidly changing with different cellular states. DNA molecules can become entangled or interlocked (catenanes) during replication, transcription or repair. If they are not relieved before cell division, it would lead to genome instability (Vos, Tretter et al. 2011). During replication, the separation of the double-helix DNA results in the over rotation of the duplex strands ahead of the moving replisome. This over winding of

the rotated duplex ahead of the fork is called supercoiling in the positive form (Figure 7). For the replisome to progress, the supercoils must be resolved. The tension resulting from the build-up of supercoils is called topological stress or superhelical stress. Several studies have shown increased topological stress as a possible cause for fork regression at stalled forks *in vivo* (Postow, Crisona et al. 2001, Postow, Ullsperger et al. 2001, Fierro-Fernandez, Hernandez et al. 2007). Notably, reversed forks are observed more frequently when the check-points are inactivated (Sogo, Lopes et al. 2002).

In 1971, a new class of enzymes called the DNA Topoisomerases which resolve superhelical tensions were discovered (Wang 1971). Since then, these enzymes have been extensively studied (Wang 2002). The function of topoisomerases can be briefly explained in three steps 1) creating a nick or a DSB by a nucleophilic attack on the DNA phosphodiester backbone by a tyrosine residue in the active site of the Topoisomerase, 2) controlled strand movements to resolve the tension, 3) religation of the strands and release of the topoisomerase by nucleophilic attack on the phosphotyrosine intermediate by the hydroxyl group of the free DNA end (Figure 8) (Champoux 2001).

Topoisomerases are generally divided into four groups IA, IB, 2A and 2B. Type I acts to resolve the supercoils by nicking one strand whereas type II severs both the strands (Champoux 2001). Within type I, IA nicks ssDNA consequently resolving the slightly unwound single strand supercoils formed in the newly replicated DNA whereas IB nicks dsDNA, therefore can resolve both the negative and positive forms of supercoiling (Champoux and Dulbecco 1972, Brown and Cozzarelli 1981, Been and Champoux 1984). Human type 2A was initially isolated from HeLa cells and its isoform type 2B was discovered a few years later (Miller, Liu et al. 1981, Austin and Fisher 1990). Type 2 enzymes bind as a dimer to the dsDNA resulting in a DSB at the close proximity of each enzyme monomer. Another double helix is allowed to pass through the gap followed by the ligation of the DSB (Nichols, DeAngelis et al. 1999).

Figure 8: Diagram showing the mechanism of three-step relaxation of supercoils by Topoisomerase I (Pommier 2010). Step I: Topoisomerase I enzyme creates a nick in the DNA. Step II: controlled rotation of the supercoil to enable relaxation. At this stage, the enzyme is bound to the DNA forming a transient TOP1 cleavage complex (TOP1cc). Step III: religation of the nicked DNA strands.

Inhibiting topoisomerases or poisoning them using chemicals result in the build-up of unconstrained supercoiling within the cells. Most topoisomerase poisons function by stabilizing the transient topo-DNA cleavable complex and as a consequence prevents the religation of the cleaved DNA. Accumulation of such complexes is recognized as replication stress by the cells as the supercoils build up halts/lags the replication fork. Especially in cells with high replication rates (pluripotent or cancer cells) where high levels of topoisomerase enzymatic activity takes place, poisoning of topoisomerases can lead to disastrous consequences (Li and Liu 2001, Romer, Nygard et al. 2013). Hence topoisomerase poisons became widely used chemotherapeutic agents.

A popular chemotherapeutic agent, which is a Topoisomerase IB (Topo IB) poison, is camptothecin (Pommier 2006). Camptothecin specifically blocks the religation of ssDNA nick created by Topo IB by trapping the end of Topo IB along with the nicked

DNA forming a Topo1 cleavable complex (Top1cc). Top1cc being an intermediate step by itself, is reversible and not considered as a damage. However, in the presence of camptothecin, interminable Top1cc leads to an increase in topological stress and if it encounters other cellular processes (such as transcription), they are converted into harmful DSBs (Horwitz, Chang et al. 1971, Tsao, Russo et al. 1993). Considering cancer cells which undergo replication frequently, more cleavage complexes are formed, resulting in increased DSBs. Furthermore, cancer cells have compromised downstream repair pathways and are more vulnerable to DNA damage (Pommier, Leo et al. 2010). However, similar to other chemotherapeutic drugs, camptothecin also affects normal cells leading to adverse side effects (Bonnetterre 1995). Interestingly, it has been demonstrated in different organisms that fork reversal occurs in 15%-40% of replication forks upon exposure to camptothecin (Ray Chaudhuri, Hashimoto et al. 2012).

1.3.2 HMGA2 and DNA topology

Direct evidence for a functional role of HMGA2 in regulating DNA topology through association with supercoils was discovered recently (Peter, Yu et al. 2016, Zhao, Peter et al. 2017). A cell based reporter system was designed, where expression of HMGA2 mediates the Renilla luciferase reporter activation driven by the HSV-TK promoter. With high-throughput screening to identify HMGA2 antagonists that can attenuate the activation of the reporter, they identified SN38, which is a highly potent metabolite of camptothecin as the antagonist. SN38 acts as a Topoisomerase 1 poison, by trapping the enzyme and nicked DNA in a complex, thereby preventing religation of the DNA strands. The study also found that SN38 did not interfere with the direct HMGA2 binding rather indirectly attenuated the reporter gene activation by increasing the efficiency of the SN38 mediated trapping of topoisomerase I in covalent complexes with DNA. It was also shown that SN38 interfered with the reporter activation in a dose dependent manner. Through *in vitro* trapping of Top1cc complexes in the presence of varying concentrations of HMGA2, it was observed that HMGA2 potentiates the SN38 mediated trapping at least by 2-to-3-fold (Peter, Yu et al. 2016). Notably, in the absence of SN38, HMGA2 by itself does not play a role in the complex formation. Experiments with mutated AT-hooks of HMGA2 protein,

imparted that all three hooks should be functionally active (in its DNA binding capacity) for this function (Peter, Yu et al. 2016).

Another study combining single-molecule analysis and AFM, discovered that HMGA2 binds to positive and negative supercoils (Zhao, Peter et al. 2017). It also revealed that at moderate levels of HMGA2, the binding of the protein with DNA occurs only at the DNA crossings of the supercoil whereas, at high levels of HMGA2, the entire supercoil becomes covered with the HMGA2 protein bridges. This data along with magnetic tweezer experiments indicate that the coating of the entire supercoil by HMGA2 scrunches and constraints the supercoils (Zhao, Peter et al. 2017). Consistent with the previous *in vitro* data (Peter, Yu et al. 2016), the analysis confirms that the AT-hooks are essential for constraintment.

1.3.3 HMGA2 and Telomeres

The first direct evidence for the presence of HMGA proteins at telomeres was the analysis of telomere proteome obtained with the technique called Proteomics of Isolated Chromatin Segments (PiCh) (Dejardin and Kingston 2009). PiCh uses specific nucleic acid probes to isolate chromatin-bound proteins, followed by Mass Spectrometric analysis of the associated proteins. With this technique they isolated the telomere-associated proteins of HeLa and identified HMGA1, but not HMGA2, presumably because HeLa cells express only HMGA1.

Another group employed the PiCh technique to study the changes in chromatin composition following telomere dysfunction. They used mouse embryonic fibroblast to deplete TRF2 to trigger telomere dysfunction, which revealed that both HMGA1 and HMGA2 were present at dysfunctional telomeres. Interestingly, only HMGA2 was pulled down in TRF2 proficient telomeres but not HMGA1. Furthermore, they could identify only HMGA1 but not HMGA2 in TRF2 deficient telomeres (Bartocci, Diedrich et al. 2014).

A recent study has shown the interaction of HMGA2 with shelterin protein TRF2 through Co-IP studies performed in cancer cells. They demonstrated that HMGA2

retains TRF2 at telomeres by interacting with TRF2 through the AT hook1-Linker1-AT hook2 region. They have also reported that silencing of HMGA2 led to reduced TRF2 binding to the telomeres as well as increase in Telomere dysfunction- Induced Foci (TIFs) suggesting telomere instability (Natarajan, Begum et al. 2016).

1.4. The CRISPR-Cas system

Using the CRISPR/cas9 tool, we generated hESC and HT1080 stable cell lines expressing GFP-FLAG or mCherry-miniSOG at the C-terminus of shelterin protein TRF1.

In order to probe various aspects of telomere biology such as spatial organization, dynamic interactions with other proteins, heterochromatin formation, subnuclear localization and replication, visualization of telomeres could become indispensable. Previously, study of telomeres has been conducted by either over-expression or knockout of the gene to understand the functions of different proteins at the telomeres. Undoubtedly, these studies have offered great insights into gene functions at multiple levels, but the consequences of altering the natural expression of proteins could have deleterious effects that cannot be overlooked. In addition, the imbalance of proteins in the cells might cause misfolding, differential localization or even unnatural posttranslational modifications. To overcome this, tagging the endogenous genes with reporters might emerge as a useful tool that could enable visualization and live cell imaging for a range of biological applications.

The recently developed genome-editing tool, CRISPR-Cas has gained enormous scientific interest owing to its affordability and feasibility in laboratories. Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated (Cas) technology derived from bacterial adaptive immune system is widely used for knock-in and knockout studies. The CRISPR locus of the host bacteria incorporates the sequence of the invading DNA called 'protospacer' to its CRISPR array locus. The CRISPR repeats are initially transcribed as part of a single long transcript and then processed by Cas proteins to form the crRNA (crRNAs), which hybridizes with the transactivating RNA (tracrRNA) forming the guide RNA (gRNA). The

gRNA forms a complex with the Cas9 nuclease. The Cas9 is directed to the cleavage site by gRNA, which is complementary to the target DNA sequence of the invading virus or plasmid. It is usually 20bp long and must be next to the Protospacer adjacent motifs called as PAM for the catalytic activity of Cas9 (Ran, Hsu et al. 2013). Upon cleavage by Cas9 at the desired locus, it is repaired either by Non-homologous End Joining (NHEJ) or Homology Directed Repair (HDR) (Figure 9). NHEJ is error prone leading to deletions, mutations or small insertions (indels) resulting in the gene knockout at the targeted site whereas in HDR, where the DSB repair is mediated by the donor template provided in the form of a plasmid or linear DNA, results in knock-ins or site-specific mutations at the site of cleavage (Sander and Joung 2014) (Figure 9).

Figure 9: Working principle of CRISPR/Cas9 endonucleases. Cas9 bound to double-stranded target DNA and induction of DSBs at a specific site can be repaired either by HDR repair mechanism or NHEJ. HDR pathway repair uses a donor template resulting in point mutations or insertions of larger DNA segments at the site of targeting. NHEJ-mediated repair produces random indels, often leading to disruption of coding region (gene knockout). Adapted from (Finotti, Breda et al. 2015).

1.5 OBJECTIVES OF THIS STUDY

1. Faithful replication of DNA till the end of chromosome is a key contributor towards genome stability. As mentioned earlier, telomeres are classified under fragile sites as mild replication stress at telomeres could result in replication stress induced DSBs. Our goal is to design a tool to visualize/track telomeres. The confined localization of telomeres at the end of chromosomes and the definite number of telomeres per cell makes the visualization studies feasible. In order to achieve this, using CRISPR/Cas9 technology, we intend to generate stable cell lines expressing GFP and FLAG or mCherry and miniSOG at the C-terminus of the shelterin protein TRF1 that directly interacts with the double stranded telomeric DNA.

2. HMGA (HMGA1 and HMGA2) proteins play an important role in the embryogenesis and differentiation but their high re-expression during tumorigenesis and its role in transforming benign tumors to malignant, classifies it as tumor-marker. Besides this, a novel role of HMGA2 during replication acting as a fork chaperone at stalled forks was established where HMGA2 (HMGA1 partially) protects the stalled fork from collapsing into DSBs by preventing the forks from endonucleolytic attack (Yu, Lim et al. 2014). Although HMGA2's affinity for binding to the branched DNA structures was attributed to this role, the precise mechanism was not known. Our goal is to establish the mechanism behind this protective function of HMGA2.

3. A recent study using single-molecule analysis found that HMGA2 binds and constrains positive and negative supercoils (Zhao, Peter et al. 2017). However, such a mechanism is yet to be validated *ex vivo*. Within cells, the relaxation of supercoils is performed by Topoisomerases. So using clinically important Topoisomerase inhibitor SN38, we aim to investigate the existence of this model, in cell-based systems. Perhaps this would provide clues to the mechanism behind the protective role of HMGA2.

In brief our objectives are

- 1) To generate HT1080 and hESC cell lines those stably express GFP-FLAG and mCherry-miniSOG endogenously at C-terminus of the shelterin protein TRF1 to visualize telomeres.
- 2) To uncover the precise mechanism of fork chaperone function of HMGA2 in HMGA2 expressing cancer cells.
- 3) To study the existence of HMGA2's supercoil constraining role on cell based systems and the association of HMGA2-Top IB using Topoisomerase I poison SN38.
- 4) To find the sites/regions in the genome which are exceptionally vulnerable to replication stress and are being protected by HMGA2 during replication stress.

2. MATERIALS AND METHODS

2.1 Endogenous tagging of *TRF1* using CRISPR/Cas9

To facilitate the study of telomeres, CRISPR/Cas9 tool was used to tag the C-terminus of *TRF1* gene with GFP-FLAG or mCherry-miniSOG. Dr. Roland Ivanyi-Nagy designed the sgRNA and generated the donor repair templates (Section 2.1.1 and 2.1.2), which were later used to engineer HT1080 and hESC.

- I performed the co-transfection of sgRNA with GFP-FLAG donor plasmid in HT1080 cells followed by puromycin selection and characterization of the picked colonies by PCR, confocal imaging, Southern blotting and Western blotting.
- Co-transfection of sgRNA with GFP-FLAG or mCherry-miniSOG donor plasmid in hESC followed by puromycin selection and generation of clonal population were carried out by Dr. Shree Harsha Vijayachandra. I characterized the selected clones by PCR, confocal imaging, Southern blotting and Western blotting.

2.1.1 Design of the sgRNA

The target sgRNA was generated using CRISPR design tool (<http://crispr.mit.edu/>). This involves identification of PAM sequence (NGG) of Cas9 ortholog (from *Streptococcus pyogenes*) at the target sequence followed by addition of 20 nucleotides upstream of PAM sequence. Three sgRNAs were selected and cloned separately into the pSpCas9 (previously known as pX330, Addgene) by two-step cloning method.

To determine the efficiency of sgRNAs a percentage of indels created by NHEJ after Cas9 cleavage, SURVEYOR assay was performed. This mainly comprises of a SURVEYOR nuclease (Transgenomic) that cleaves both DNA strands to any mismatches. To carry out the assay, the target DNA was amplified denatured and then

allowed to hybridize as heteroduplex. The hybridized complex was treated with SURVEYOR nuclease and analyzed by agarose gel electrophoresis (Figure 10). The band intensity of each fragment was quantified with the ImageJ software. The percentage of indels was calculated with the formula derived by (Ran, Hsu et al. 2013).

$$\text{Indel (\%)} = 100 \times (1 - (1 - f_{\text{cut}}))$$

$$\text{Where, } f_{\text{cut}} = (b + c) / (a + b + c),$$

a is the intensity of the undigested PCR product and b and c are the intensities of each cleavage product. Based on the SURVEYOR assay results, a single guide RNA was chosen.

Figure 10: Graphical representation of steps involved in sgRNA selection. Three sgRNAs per cell line were chosen *in silico* using CRISPR design tool (<http://crispr.mit.edu/>). The selected sgRNAs were cloned separately into the pSpCas9 vector and transfected into cells. SURVEYOR assay was used to measure the ability of each sgRNA to facilitate targeted cleavage.

2.1.2 Design of the repair template

The donor templates (repair templates) were designed using the Gibson assembly (New England Biolabs). The template cassette is flanked on either side by 1000bp sequence that is homologous to the *TRF1* gene and are termed as the Right and Left homology arms (Figure 11, 12). The homologous sequence also has a mutated version of the PAM sequence that replaces the existing PAM during repair. Hence if a site is repaired once it cannot be used again to form a complex with Cas9. The template also has loxP sites to remove the selection cassette.

Figure 11: Donor plasmid of GFP-FLAG template. This vector was used for constructing the GFP-FLAG expressing HT1080 and human ES cells in this study. The left and right arm regions are homologous to the TRF1 gene.

Figure 12: Donor plasmid of mCherry-miniSOG template. This vector was used for constructing the mCherry-miniSOG expressing human ES cells in this study. The left and right arm regions are homologous to the TRF1 gene.

2.1.3 Co-transfection and selection of positive clones

The donor template and the pSpCas9 were co-transfected into HT1080 or hESC cell lines. After transfection, cells were kept under puromycin selection for two weeks and clonal populations were obtained (Figure 13). Each clone was verified for the integration at both right and left homology arm junctions by PCR. The positive clones were plated on glass bottom plate (Nunc, Thermo scientific) and viewed under confocal microscope (Corrsight, FEI) for fluorescence signal (Figure 13). The clones that were positive from both PCR and confocal microscopy analysis were expanded.

Figure 13: Graphical representation of the steps involved in CRISPR/Cas9 knock-in strategy performed in the study. (A, B): The pSpCas9 vector bearing sgRNA (A) to cleave at the targeted locus was co-transfected with the donor template (B) containing either GFP-FLAG or mCherry-miniSOG to perform the homology directed repair at the cleaved site. (C): The transfected cells were subjected to puromycin selection and subsequently clonal cells were generated. (D): The clones were analyzed under confocal microscope for expression of GFP/mCherry. (E): The fluorescent clones were annotated and verified for the successful integration of homology arms by PCR and sequencing. (F): Clones those were positive were expanded for further characterization.

2.1.4 Characterization of the selected clones

The targeted locus of the selected clones was verified by sequencing. In addition, Western and Southern blots were performed to check for the presence of tagged protein and to verify single-copy integration, respectively (Figure 13).

2.2 Cell culture

The cell lines used in this study are as follows.

- **Human embryonic stem cell (hESC)** line ‘Genea 047’ obtained from Genea Biocells, Sydney, Australia were cultured in Genea M2 media (Genea Biocells) supplemented with Penicillin and Streptomycin at 25 Units/ml each (Gibco, Life technologies). The culture conditions were 37°C under 5% CO₂ and 5% O₂. The cells were grown on Collagen I coated cell culture dishes (Biocoat, Corning). For passaging, adherent cells were rinsed with 1xPBS and detached using the passaging solution (Genea Biocells) by incubating for 4-6 minutes at 37°C. The cells are collected with 5mL of Neutralization solution (Genea Biocells) and centrifuged at 500g for 5mins at 4°C. After removing the neutralization solution, cells are resuspended in M2 medium and plated in new plates in 1:4 ratio. For selection of recombinants and maintenance of targeted clones, Puromycin (300 ng/ml) (Gibco, Life technologies) was included in the growth medium.
- **HT1080 and Recombinant HT1080-C1 and HT1080-C2** (human origin fibrosarcoma) Recombinant HT1080-C1 and HT1080-C2 cell lines were generated previously in our lab (Yu, Lim et al. 2014). Individual colonies were picked after transduction of HT1080 cells with pTRIPZ-shHMGA2 (Thermo Scientific). HMGA2 knockdown was achieved by inducing shHMGA2 with 2µg/mL Doxycycline once a day for four days. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (GIBCO, Life Technologies) supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine and 100 Units/ml of penicillin and streptomycin each (Gibco, Life technologies). For selection of puromycin-resistant clones, puromycin (Gibco, Life technologies) was added in the growth medium (1 µg/ml). Passaging was done by rinsing the plate with 1xPBS followed by Trypsinization at 37°C for 5mins.
- **HMGA2 expressing A549 cells** (human origin lung epithelial sarcoma) were previously generated in our lab (Summer, Li et al. 2009) by electroporating pEF1-hHMGA2-His-Neo into A549 cells followed by Neomycin selection. Three clones namely 1.3, 1.5 and 1.6 were chosen based on western blot analysis for HMGA2 expression. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (GIBCO, Life Technologies) supplemented with 10% fetal bovine

serum (FBS), 1% L-glutamine and 100 Units/ml of penicillin and streptomycin each (Gibco, Life technologies). Passaging was done by rinsing the plate with 1xPBS followed by Trypsinization at 37°C for 5mins.

- **HMGA2 expressing HeLa** (Cervical epithelial adenocarcinoma) cells were generated previously generated on our lab by lipofectamine (Invitrogen) transfection of pEF1-hHMGA2-His-Neo into HeLa cells followed by Neomycin selection. Three clones namely P2 P8 and P9 were chosen based on western blot analysis for HMGA2 expression (Summer, Li et al. 2009). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (GIBCO, Life Technologies) supplemented with 10% fetal bovine serum (FBS), 1% L-glutamine and 100 Units/ml of penicillin and streptomycin each (Gibco, Life technologies). Passaging was done by rinsing the plate with 1xPBS followed by Trypsinization at 37°C for 5mins.
- **H1299** (Human origin non small cell lung carcinoma) H1299^{Parental} and HMGA2 knockout (KO) H1299^{HMGA2-KO} cells were generated using CRISPR/Cas9 were a gift from our collaborator at Novartis, La Jolla. The knockout of HMGA2 cells was confirmed by western blot analysis by the collaborator. The frameshift of the gene was confirmed with PCR and sequence analysis in our lab. The cells were cultured in RPMI media (GIBCO, Life Technologies) supplemented with 10% fetal bovine serum (FBS), 100 Units/ml of penicillin and streptomycin each (Gibco, Life technologies). Passaging was done by rinsing the plate with 1xPBS followed by Trypsinization at 37°C for 7mins.

2.3 Genomic DNA Extraction

- **Method 1:** Genomic DNA was isolated using the DNeasy Tissue kit (Qiagen) following their standard protocol. Briefly pellets from 1×10^6 cells were resuspended in PBS containing Proteinase K and RNase A. Suspended cells were lysed by adding AL buffer and incubating at 56deg for 10min. After addition of absolute ethanol, the suspension was transferred to spin columns provided in the kit and washed by centrifugation with AW1 and AW2 buffer once each. The genomic DNA was eluted using buffer AE into an eppendorf and quantified using nanodrop. The quality of DNA was checked by agarose gel electrophoresis.

- **Method 2:** Cells were trypsinized, washed once with PBS and resuspended in lysis buffer (100mM NaCl; 20mM Tris pH8; 10mM EDTA; 0.5% SDS and 400µg /mL Proteinase K). The lysate was incubated overnight at 55°C. Saturated NaCl (6M) was added to the lysate and mixed by inverting the tubes and incubated on ice for 10min. The mix was centrifuged at 13,000rpm for 15min at 4°C and the supernatant was collected in a fresh tube kept on ice. Cold isopropanol was added to the supernatant and mixed until the DNA threads became visible. The tubes were then incubated on ice for 15min and centrifuged. The pellet was washed with ethanol and resuspended in TE buffer.
- **Method 3:** Cells were harvested by trypsinization, centrifuged and washed with PBS. The pellet was resuspended with Quick Extract Reagent (Epicenter) and mixed by vortexing for 15sec. The sample was heated for 6min at 65°C, mixed by vortexing for 15sec followed by boiling for 2min at 98°C.

2.4 Agarose gel electrophoresis

Based on the size of DNA fragments to be separated, appropriate amount of agarose (LONZA) was solubilized in TBE buffer (100mM Tris, 100mM Boric Acid, 2mM EDTA, pH 8.3) by heating in a microwave oven. 1 mg/mL of ethidium bromide was mixed in and let to solidify. The samples were mixed with 6x Loading Dye (NEB) and loaded in an electrophoresis apparatus containing TBE buffer. The DNA was visualized under UV.

2.5 Pulsed-Field Gel Electrophoresis (PFGE)

Cells were seeded such that they were 60-70% confluent at the time of drug treatment. Following treatment, cells including the detached population were harvested and embedded into 2% low melting agarose plugs with disposable plug model (Bio-Rad). After solidification, plugs were incubated in lysis buffer (100 mM EDTA, 0.2% (w/v) sodium deoxycholate, 1% (w/v) sodium lauryl sarcosine and 1 mg/ml proteinase K; Promega) at 50°C for 24h and washed 4x in TE buffer for 1h each. After running the gel (1% chromosomal-grade agarose; Bio-Rad) on CHEF DR II equipment (Bio-Rad)

in 1x TAE buffer with parameters 120 field angle, 5-30 s switch time, 4 V/cm and 14 °C for 17h, visualization was done by staining the gel with ethidium bromide.

2.6 Terminal Restriction fragment (TRF) analysis

Genomic DNA was purified as described in Section 2.3; Method 2. Briefly, 2µg gDNA isolated from the cells were digested with 20U of each HinfI and RsaI overnight at 37°C. The digested DNA was separated on 0.8% gel 30V for 16h. Southern blotting employing the telomere probe, was performed using TeloTAGGG Telomere Length Assay kit (#12209136001, ROCHE) as per the manufacturers' protocol. To detect subtelomeres, undigested genomic DNA was separated using either 0.8% agarose gel electrophoresis or PFGE and processed the same way as mentioned above.

2.7 Southern blotting

Genomic DNA was purified as mentioned in Section 2.3; Method 1. 15 µg of DNA was subjected to restriction digestion using 50 U of EcoRI/PstI enzyme overnight at 37°C. DNA was ethanol precipitated and dissolved in 20 µL TE buffer (pH 8.0). Targeting vector was linearized with EcoRI/PstI restriction enzyme and diluted to 10^7 , 10^8 copies per µL. Digested gDNA samples were resolved overnight on a 1% agarose gel in 1x TAE (Tris-Acetate-Boric acid) buffer, with 1 kb DNA ladder (New England Biolabs). The probes were generated using the DIG-High Prime DNA Labeling and Detection Starter Kit II (Roche) according to the manufacturers' protocol. The probe-target hybrids on the blots were detected by chemiluminescent assay followed by exposure to an X-ray film (Kodak MXG film, Kodak) and developed on a Kodak X-OMAT 2000 Processor.

2.8 PCR analysis

PCR was performed using GoTaq Flexi DNA polymerase (Promega). The DNA from cells was either extracted using Section 2.3; Method 1 or 3 and used as template. The PCR conditions used are stated below.

Initial Denaturation	95°C for 5 min	
Denaturation	95°C for 30 sec	
Touchdown-Annealing	61° - 51°C for 30 sec	10 cycles; -1°C/cycle
Touchdown-Extension	72°C; 1min/kb	
Denaturation	95°C for 30 sec	25 cycles
Annealing	51°C for 30 sec	
Extension	72°C; 1min/kb	
Final extension	72°C for 10 min	

Table 1: List of primers (5'-3') used in this study

Puro FW	GAGCTGCAAGAACTCTTCCTCAC
Puro RV	GCTCGTAGAAGGGGAGGTTG
GFP FW	CCACATGAAGCAGCACGACTTC
GFP RV	GGGTGTTCTGCTGGTAGTGGTC
P12	TGCTGGGAGAAAATTCCTAAAC
P21	GTCCCTGCTCTCGACTCTTT
P22	GCTCATGATGCCAGCCTTG
P16	CCATGGGTACTCAGGTATTAC
P41	GGTTACGTTACAGGGTTGAGGT
P55	GTCCCTCCTTGAAGTCGATGC
P86	GACTACAAAGACCATGACGGT
P88	ATGACCGAGTACAAGCCCAC
P90	CGAGATCCACTAGTTCTAGC
P91	GCTAGAACTAGTGGATCTCG
P73	CTCAGAGGCAGTGGTACCTCT
P40	TGTTTAGAGAAGTCTTCCGTCTTT
P04	GCAGGTGTTTGAATGAGGT
P72	AGAGGTACCACTGCCTCTGAG
Puro 24	CACCGTGGGCTTGTACTCGGTC
PRmC	GCTGGAGAATTTCTTGTAAGTCTAG
PGK RV	CAGGTGAATATCAAATCCTCCTCG

HMGA2 FW	CAACTCCTGATCCCAACCCG
HMGA2 RV	GCCGTGCACAATAGCGAAAG

2.9 Sequencing of the PCR products

After PCR amplification, the products were analyzed on agarose gel electrophoresis and stained with ethidium bromide for visualization. The desired band was excised with a scalpel and transferred to an eppendorf. The DNA was extracted from the agarose piece using QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's protocol. The extracted DNA was then quantified using nanodrop and sequenced with respective primers using 1st Base DNA sequencing service. The sequencing data obtained was analyzed in Vector NTI software (Thermo Fisher Scientific).

2.10 TA-cloning

After agarose gel electrophoresis analysis of the PCR product, the gel was stained with ethidium bromide for visualization. The desired band was excised and purified using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's protocol. The purified PCR product was cloned into pGEM®-T Vector System following the manufacturer's protocol. The resulting reaction mix was transformed into DH5α competent cells and plated on an agar plate containing ampicillin. Following day, few colonies were picked and amplified by colony PCR. The resulting PCR product was purified using the QIAquick PCR Purification kit and sequenced.

2.11 Immunoblotting

Cells were harvested, washed once with PBS and resuspended in ice-cold RIPA buffer (150mM Sodium Chloride, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, 50mM Tris, pH 8) containing Protease Inhibitor Cocktail (Roche). The lysate was sonicated on ice for 10x, 3sec each at 5W. The sheared lysates were centrifuged at 13,000rpm for 10min at 4°C. The supernatant was separated and the protein concentration was determined using Protein Assay Dye Concentrate (Bio-Rad) using

bovine serum albumin (BSA) as a standard. 30 to 40µg of protein per sample was analyzed on SDS-PAGE or gradient precast gels according to the molecular weight of the protein of interest and transferred to PVDF membranes with 0.2µm pore size (Bio-Rad). To prevent non-specific binding, membrane was blocked with 5% nonfat milk (Bio-Rad) or using Superblock Blocking Buffer (Thermo Fischer Scientific) and incubated with primary antibody in blocking buffer overnight at 4°C followed by secondary antibody incubation for 1hr at RT. The immunoreactive bands were detected using the Western HRP substrate (Luminata Forte, Millipore) or EMD Millipore Immobilon Western Chemiluminescent HRP Substrate (ECL) in an infrared Imager (LAS-4000, Fuji). β -Actin was used as internal control. Band intensities were quantified using ImageJ.

Table 2: List of antibodies used in this study

Anti-TRF1	Pierce (PA5-28423)	Rabbit polyclonal (1:1000)
Anti-FLAG	Sigma (F7425)	Rabbit polyclonal (1:1000)
Anti-HMGA2	CST (#5269S)	Rabbit polyclonal (1:1000)
Anti- β -actin	Sigma (A2228)	Mouse monoclonal (1:5000)
Anti-Topoisomerase I	Abcam (Ab109374)	Rabbit monoclonal (1:2000)
Anti-Topoisomerase II alpha	Topogen (TG2011-1)	Rabbit polyclonal (1:2000)
Anti-Topoisomerase II alpha + Topoisomerase II Beta	Abcam (Ab109524)	Rabbit monoclonal (1:2000)
Anti-HMGA1	CST (#7777)	Rabbit monoclonal (1:1000)
Anti-TRF1	Bio-rad (#VMA00032)	Mouse monoclonal (1:1000)

Anti-TRF2	Santa Cruz (sc-9143)	Rabbit polyclonal (1:500)
Anti-rabbit (Secondary)	Dako (P 0448)	Goat anti-rabbit (1:10,000)
Anti-mouse (Secondary)	Dako (P 0260)	Rabbit anti-mouse (1:10,000)
Anti-rabbit (secondary) Alexa Fluor 555	Life Technologies (A31572)	Donkey anti-rabbit (1:200)

2.12 Cytotoxicity assays:

2.12.1 Cell survival test: Cells were seeded as triplicates in black walled 96-well plate and let to attach overnight. Different doses of drugs were mixed to the media and added on the plate. Media was changed every day during the course of treatment for both treated and untreated cells. Following the treatment, cell viability was determined using the Cell Counting Kit-8 (CCK8) (Enzo Life Sciences) as per manufacturer's protocol. Briefly, the CCK8 solution from the kit was diluted with the media and added to the wells. The plate was incubated for 3h and O.D. was measured at 450nm in a plate reader (TECAN Infinite M200 Pro).

2.12.2 Caspase assay: Cells were seeded as triplicates overnight in a white walled 96-well plate and the following day, treated with different doses of SN38 for different period of time as indicated. Following treatment, caspase activity was measured by adding 100µl of room temperature equilibrated Caspase/Glo® 3/7 reagent to each well and incubated for 1h at room temperature. Luminescence was measured using a plate reader (TECAN Infinite M200 Pro).

2.13 Immunostaining

For co-localization studies between TRF1 and TRF2, HT1080 *TRF1*-GFP-FLAG targeted clones grown on coverslips were washed once with PBS, permeabilized with 0.5 % Triton in pH 7.4 PBS for 10 min and washed twice with PBS for 5 min to remove soluble proteins. Cells were fixed in 3% paraformaldehyde in PBS (pH 7.4) for 15 min, washed with PBS 3x. Blocking was performed with 5% BSA for 1h.

Staining with primary antibody followed by the secondary antibody incubation was performed in PBST with 1% BSA for 1hr each. The coverslips were mounted on the glass slide with the aid of mounting medium and sealed with clear-coat polish.

2.14 Optogenetic oxidative stress induction

Human ESC derived clone A3, which has the C-terminus of *TRF1* tagged with mCherry-miniSOG was used for the experiment. WT hESC cells served as control. Increasing number of cells were plated in three wells for different conditions in glass bottomed, black walled collagen coated 96-well plates. After determining the intensity of the light through preliminary experiment, each well containing cells were exposed to blue light for different times as indicated. After the exposure, plates were left in the incubator for 24h before changing media. The cells were let to grow in the incubator for 5 days and media was changed as required. Cell viability assay was performed using the CCK8 following the manufacture's protocol.

2.15 Next Generation Sequencing Analysis

10^6 H1299^{Parental} and H1299^{HMGA2-KO} cells were seeded in eight 10cm dishes each and let to attach overnight. Fresh media containing 2 μ M SN38 was added to the plate and treated for 48h. Following the treatment, the cells were harvested and processed for PFGE as described in previous section 2.5 The PFGE was run in a CHEF DR II equipment (Bio-Rad) in 1x TAE buffer with parameters 120 field angle, 5-30 s switch time, 4 V/cm and 14 °C for 14h. The gel was stained with ethidium bromide for visualization under the UV light. Lambda PFG marker (NEB) was run alongside samples as the standard.

The smear ranging from 30kb to 150kb was excised. Gel slices from eight lanes were pooled for each sample to obtain enough DNA for deep sequencing. DNA was extracted using the Zymoclean Large Fragment DNA recovery kit (Zymo Research) following the manufacturer's protocol. The quality of DNA was checked by agarose gel electrophoresis. 1.5 μ g DNA was sent for next generation sequencing. As controls, 1.5ug genomic DNA extracted as per in section 2.3; Method 1 from H1299^{Parental} and H1299^{HMGA2-KO} (untreated) cells were also sequenced.

High-throughput sequencing was performed using Illumina HiSeq High Output 2x101bp (multiplexed). After the addition of Illumina adapters, 6 cycle PCR was performed for library preparation. Each library was sequenced to a depth of at least 300 million reads. The sequenced reads were checked for quality using <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. The adapters were trimmed using Trimmomatic (Bolger, Lohse et al. 2014). The sequences were aligned to GRCh38 using bowtie2, and paired-end mode was used with default parameters (Langmead, Trapnell et al. 2009). The resulting BAM files were binned using the bamCoverage tool using the galaxy server (www.usegalaxy.org). Since our fragments were large ranging from 30-150kb, we started with 20kb binning to determine the enrichment in the SN38 induced DSB fraction. Enrichment fold changes were computed against the untreated controls for each sample H1299^{Parental} and H1299^{HMG2-KO} respectively. FDR was calculated using the FDRtool (Strimmer 2008) and regions under 1% FDR were considered enriched or depleted. UCSC genome browser was used to visualize the coverage for individual tracks and genome-wide distribution was plotted using Circos (Krzywinski, Schein et al. 2009).

3. RESULTS

Chapter 1

3.1 CRISPR/Cas9 mediated tagging of *TRF1* in HT1080 and hESCs

In order to probe into various aspects of telomere biology such as subnuclear localization, heterochromatin formation and the dynamic behavior of telomeric DNA during replication and induced replication stress, we targeted the gene *TRF1* (encoding shelterin protein TRF1) with two sets of reporter systems. One of the reporter systems includes GFP and FLAG tag, which could enable visualization of the telomeres as well as help in affinity pull-down assays to study the physical interactions between proteins and validating protein-protein interactions. The other reporter includes mCherry for visualization along with miniSOG (mini Singlet Oxygen Generator) green fluorescent flavoprotein that can be detected by correlated light and electron microscopy (EM) (Shu, Lev-Ram et al. 2011).

We established a workflow (Figure 13) to create stable cell lines expressing GFP-FLAG and mCherry-miniSOG fused to endogenous C-terminus of *TRF1* using CRISPR/Cas9 system. For this purpose, we designed sgRNAs and a donor template bearing homology arms of about 1000bp to facilitate the integration of expression cassette at 3' end of last exon of *TRF1* gene at chromosome 8. This resulted in a fusion of a GFP/miniSOG cassette to endogenous *TRF1*, thus maintaining the native level of fusion protein expression.

The details of CRISPR/Cas9 targeting has been explained below:

3.1.1 Selection of the sgRNA

Using the online CRISPR Design tool, three sgRNAs were generated to target the C-terminus of the *TRF1* gene (crispr.mit.edu).

Figure 14: Schematic representation of the different sgRNAs chosen *in silico*.

(sg01, sg02 and sg03 (indicated in blue); selected by the online CRISPR Design tool) along with PAM sequences (indicated in red), targeting exon 10 of *TRF1* genomic locus. The stop codon TGA of *TRF1* gene is also indicated.

Surveyor assay was performed with all the three sgRNAs to ascertain the percentage of indels created by NHEJ after Cas9 cleavage (Section 2.1.1) As a control; pSpCas9 vector without sgRNA was included. The intensity of the bands was analyzed with ImageJ software and the indel chart (Table 3) was created based on a formula mentioned in the Materials and Methods (Section 2.1.1).

Figure 15: SURVEYOR assay to measure the ability of each sgRNA to facilitate targeted cleavage. A gel picture showing the fragments generated by three different sgRNAs in HT1080 cells using Surveyor assay. Sg01: sgRNA1, sg02: sgRNA2, sg03: sgRNA, pSpCas9: control plasmid without any sgRNA, M: 100bp ladder. Sg01 (indicated in the box) was chosen for generating the knock in cell line.

Table 3: A table showing the intensity of each band and the percentage of indels created with different sgRNAs

sgRNA	Uncut	Upper band	Lower band	Indel %
Sg01	12527	894	578	5.4
Sg02	11409	559	68	2.64
Sg03	9401	1171	315	7.07

Although both sg01 and sg03 seem to work better than sg02 (Figure 15, Table 3), sg01 was chosen for the transfection, due to its close proximity to the stop codon. (Contributed by Dr. Roland Ivanyi-Nagy).

3.1.2 Design of the repair template for knock-in

The repair template for knock-in, also called as the donor plasmid was assembled using Gibson assembly (Contributed by Dr. Roland Ivanyi-Nagy) (Figure 16). To select the successful integrants, puromycin resistant gene (selection marker) was included in the donor plasmid.

Figure 16: Schematic diagram of the GFP-FLAG and mCherry-miniSOG donor plasmid construct used for knock in at *TRF1* locus. Donor plasmid to deliver GFP-FLAG (top panel) and mCherry-miniSOG (bottom panel) used in this study to deliver GFP-FLAG or mCherry-miniSOG gene at the TRF1 locus.

3.1.3 Co-transfection of the sgRNA/Cas9 vector and GFP-FLAG donor plasmid in HT1080 cells

HT1080 cells were co-transfected with the donor plasmid along with the pSpCas9-sgRNA01. As a control, donor plasmid along with the pSpCas9 without the sgRNA was used. Initially 250ng or 500ng of both donor and pSpCas9 were co-transfected and the cells were selected under 1 μ g/ml of puromycin. After 10 days of selection, there was negligible cell death indicating that the selection efficiency was low, which was believed to be because of high random integration and off target effects. In order to overcome this, the amount of donor plasmid was reduced (Sander and Joung 2014). The transfection was repeated with 60ng of donor plasmid and 500ng of pSpCas9-sgRNA to have a stringent selection and in a hope to surpass the random integration issue. After 2 weeks of selection under puromycin, there was significant cell death indicating improvement in selection efficiency. These puromycin resistant cells were expanded for confocal microscopy imaging for GFP expression (Figure 17) and eventually to generate clonal population.

Figure 17: Confocal image analysis of the bulk cells expressing GFP-TRF1. Following co-transfection of the sgRNA and donor plasmid, puromycin selection was performed for two weeks; the bulk population of cells expressing GFP was imaged using a confocal microscopy.

3.1.4 Generation of clonal population from the targeted HT1080 cells

Single cell FACS sorting of GFP positive cells for the generation of clonal population was not effective due to low intensity of signal that was only detectable by a confocal microscope. Therefore, in order to obtain clonal population, targeted cells were plated in very low confluency of about 1000-1500 cells per 10 cm dish. Each cell was

allowed to form a colony, which was later picked and grown in 96 well plate. In total, 55 clones were picked for subsequent screening.

3.1.5 Imaging and PCR screening of the targeted HT1080 clones

In order to determine the successful integration of left and right homology arms of the donor in *TRF1* locus, gDNA was obtained from the 55 colonies using QuickExtract (Section 2.3; Method 3) and PCR amplification was carried out with specific primers (P21 and P55) for left homology arm integration (Figure 18).

Figure 18: PCR amplification of the Left Homology arm integration (LHA) in HT1080 clones.

Primer pair used for the amplification is indicated in the schematic. The clones highlighted A2, A8, A11, D11 and E11 were sequenced and chosen for further analysis. Each lane has gDNA from puromycin resistant clones obtained after transfection of donor plasmid along with Cas9 expressing plasmid, WT; gDNA from wild type HT1080, PC; Positive control, WC; Water control, bulk; bulk gDNA obtained after transfection of donor plasmid along with Cas9 expressing plasmid, M; 1kb ladder.

This resulted in 5 clones (HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#E11}) being positive for left homology arm (Figure 18).

PCR for integration of right homology arm in *TRF1* locus using gDNA from all the 5 clones with primers (P90 and P22) showed positive integration in each (Figure 19).

Figure 19: PCR amplification of the Right Homology arm (RHA) integration in HT1080 clones.

All five clones tested show positive amplification of the RHA. Primer pair used for the amplification is indicated in the schematic. Lanes A2, A8, A11, D11, E11 gDNA from puromycin resistant clones HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#E11} obtained after transfection of donor plasmid along with Cas9 expressing plasmid, WT; gDNA from wild type HT1080, WC; Water control, bulk; bulk gDNA obtained after transfection of donor plasmid along with Cas9 expressing plasmid, M; 1kb ladder.

Positive clones from both PCR analyses were plated in a glass bottom plate and confocal microscopy analysis showed that all the 5 clones were positive for GFP expression localizing to the nuclear foci (telomeres) (Figure 20).

Figure 20: Confocal microscopy image of HT1080 clones HT1080^{G#A2} (left panel) and HT1080^{G#D11} (right panel) showing GFP expression at TRF1 locus.

Next the integration of whole cassette at *TRF1* locus was confirmed by PCR amplification (Figure 21) of different regions of the donor plasmid with various overlapping primer pairs (Figure 21a).

Figure 21: PCR amplification and sequence verification of the donor plasmid integration in HT1080 clones. The desired band from each PCR sequenced, is highlighted in the box.

a) Schematic representation of the donor plasmid at the targeted genomic locus and the position of primer pairs indicating the regions amplified for sequencing. b) PCR amplification of the primer pair P86-P91; c) PCR amplification of the primer pair P88-P41; d) PCR amplification of the primer pair P88-P16. Lanes A2, A8, A11, D11, E11 are gDNA from the clones HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#} respectively, bulk; gDNA from bulk cells after puromycin selection, M; 1kb ladder, WC; water control.

3.1.6 Screening for potential bi-allelic gene targeting event in the tagged HT1080 clones

In order to check if the clones had monoallelic or bi-allelic targeting, PCR analysis was done using the forward primer at the genomic region and the reverse primer at the 3'-UTR region. Since the donor plasmid was designed such that it has a 3'-UTR region between the loxP and puromycin cassette, the bi-allelic targeted clones would result in 1.5kb band. On the other hand, in the case of monoallelic targeting, the PCR results in a band corresponding to the size of WT, 0.6kb (Figure 22).

Figure 22: Schematic representation of the PCR amplification to screen for monoallelic or bi-allelic targeting events. Primers P40 and P41 located at the homology arms are indicated. The three possible outcomes of targeting in agarose gel analysis and their corresponding PCR amplifications are depicted.

The PCR performed on the five positive clones with primers at genomic region (P40-P41) resulted in a band, indicating that all the clones had only one allele targeted (Figure 23a). The PCR product obtained with the untargeted allele was extracted and sequenced which showed an insertion in the untargeted allele in all the 5 clones (Figure 23b), resulting in a frame shift. Hence the C-terminus end of the protein at the untargeted allele has additional 8 amino acids (Figure 23c).

Figure 23: PCR amplification of the untargeted allele in HT1080 clones and the sequencing outcome.

- a) PCR amplification to screen for mono/bi-allelic targeting of the selected clones showing monoallelic targeting. Lanes A2, A8, A11, D11, E11; gDNA from the clones HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#E11} respectively, WT; gDNA from wild type HT1080, WC; Water control, M (kb); 1kb ladder, M (bp); 100bp ladder,
- b) Sequencing of the untargeted allele PCR product indicating an indel upstream of the stop codon of *TRF1* gene in the clones as compared to the WT *TRF1* sequence.
- c) TRF1 protein sequence of the untargeted allele of the clones showing an additional 8 amino acids as a result of frameshift as compared to the WT TRF1 protein sequence.

3.1.7 Southern blot analysis of the integrated HT1080 clones

In order to verify the copy number of the donor plasmid integrated in the positive clones (HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#}), southern blots were performed (Figure 24b, 24c). Genomic DNA from each clone was extracted as per Section 2.3; Method 1 and subjected to restriction digestion with EcoRI enzyme. PCR-derived digoxigenin-labelled probe complementary to the GFP and puromycin gene were used to detect left and right homology arms integration respectively. This resulted in an expected 3kb fragment for left junction and 4kb

fragment for the right junction indicating single copy integration in all the clones (Figure 24).

Figure 24: Southern blot analysis of targeted HT1080 clones with internal GFP and puromycin probe.

a) Representation of the *TRFI* genomic locus with integrated plasmid showing *EcoRI* restriction sites and expected band size for the left junction with a GFP probe and the right junction with a puromycin probe in Southern blot,

b) Southern blot analysis of targeted HT1080 clones with probe complementary to the GFP gene showing the expected band size of 3kb as highlighted in the box.

c) Southern blot analysis of targeted HT1080 clones with probe complementary to the puromycin gene showing the expected band size of 4kb as highlighted in the box.

Genomic DNA of clones (HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#E11}) was digested with *EcoRI* and southern blot was performed with probes as mentioned. Lane 10⁷; 10⁷ copies of plasmid, lane 10⁸; 10⁸ copies of plasmid, lanes A2, A8, A11, D11, E11 are gDNA isolated from clones HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#E11} respectively, lane WT; gDNA from wild type HT1080 cells, M; 1kb ladder.

3.1.8 Immunoblotting of the tagged TRF1 protein in HT1080 clones

In order to probe the tagged as well as endogenous TRF1 protein, western analysis with anti-TRF1 antibody was carried out. This resulted in the expected size of 55 KDa for untagged endogenous TRF1 protein and 100 KDa for the tagged fusion TRF1 protein (Figure 25).

Figure 25: Western Blot analysis to detect the endogenous as well as fusion TRF1 in targeted HT1080 clones probed with anti-TRF1 antibody. The GFP-FLAG-TRF1 fusion protein is detected at 100kb and the endogenous TRF1 at 55kb. Lane WT; whole cell lysate of wild type HT1080 cells, lanes A2, A8, A11, D11, E11; whole cell lysates of the targeted clones HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#E11} respectively, M; molecular weight-size marker.

Although the western blot with anti-TRF1 antibody shows a convincing band for the fusion protein, various TRF1 antibodies that had been tested in the laboratory were inconsistent and often resulted in non-specific bands. So, western blot analysis was done to probe FLAG and GFP in the fusion-TRF1 protein with anti-FLAG and anti-GFP antibodies. The expression of the tagged protein was confirmed with the anti-FLAG antibody (Figure 26). Clone HT1080^{G#E11} showed extra bands while clones HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11} and HT1080^{G#D11} had only the desired fusion protein band (Figure 26). However, the immunoblot with anti-GFP antibody was not conclusive after repeated attempts with different antibodies due to high non-specific bands (data not shown).

Figure 26: Western Blot analysis for fusion TRF1 in targeted HT1080 clones probed with anti-FLAG antibody. The GFP-FLAG-TRF1 fusion protein is detected at 100 kb as indicated by the arrow. NC1; whole cell lysate of WT hESCs, NC2; whole cell lysate of WT HT1080 cells, PC; whole cell lysate of GFP-FLAG tagged hESCs clone included as a positive control, lanes A2, A8, A11, D11, E11 are whole cell lysates of the targeted clones HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11}, HT1080^{G#D11}, HT1080^{G#E11} respectively, M; molecular weight-size marker. β -actin was used as loading control.

Based on the characterization of the five clones, clone HT1080^{G#E11} was eliminated while HT1080^{G#A2}, HT1080^{G#A8}, HT1080^{G#A11} and HT1080^{G#D11} were expanded for future studies.

3.1.9 Co-localization study with the derived HT1080 clones

Co-localization of stably tagged TRF1 with endogenous TRF2 was performed using clones HT1080^{G#A11} and HT1080^{G#D11} (Figure 27). The immunostaining of endogenous TRF2 with red fluorescent protein and GFP foci of TRF1 protein were seen co-localizing in some cells as expected (Figure 27), confirming the stable expression at the targeted TRF1 protein. However, in a fraction of cells the co-localization is not distinct possibly due to sub-optimal immunostaining of TRF2. The quantification of the co-localization using JACoP in ImageJ was not conclusive.

Figure 27: Expression of GFP-FLAG tagged TRF1 (green), immunostaining of the endogenous TRF2 (red) and the co-localization of TRF1 and TRF2 foci in HT1080 clones. The images were captured in grey-scale and colorized later.

3.1.10 Endogenous tagging of hESCs *TRF1* with GFP-FLAG and mCherry-miniSOG using CRISPR/Cas9:

Stable cell lines tagging the C-terminus of *TRF1* of hESCs with GFP-FLAG and mCherry-miniSOG with puromycin cassette as selection marker was generated in our lab (by Dr. Shree Harsha) (Figure 28a, b). Among many positive clones, the characterization of four clones namely hESC^{G41}, hESC^{G46}, hESC^{G55} and hESC^{C17} (G representing GFP tagging and C representing mCherry tagging) by PCR, Confocal imaging, Southern blotting and Western blotting were performed.

Figure 28: Schematic diagram of a) GFP-FLAG donor plasmid construct and b) mCherry-miniSOG donor plasmid construct used in this study.

3.1.11 Imaging and PCR screening of the targeted hESC clones

Imaging of the clones was performed using confocal microscopy to confirm the GFP or mCherry fluorescent protein expression (Figure 29 for GFP and Figure 30 for mCherry). In order to determine the successful integration of left and right homology arms in *TRF1* locus, gDNA was obtained from all four clones (Section 2.3; Method 1). PCR amplification using the specific primers (P21 and P55 for GFP; P21 and P73 for mCherry) confirmed the left homology arm integration in all the 4 clones (Figure 31).

Figure 29: Confocal microscopy images showing GFP signal for hESCs clones hESC^{G41}, hESC^{G46}. Corresponding DAPI stain images are shown.

Figure 30: Confocal microscopy images showing mCherry signal for clone hESC^{C17}. Corresponding DAPI stain image is shown.

Figure 31: PCR amplification of the Left Homology arm (LHA) integration in hESC clones.

Primer pairs P21-P55 used for the amplification of the GFP-FLAG tagged clones and P21-P73 for the mCherry-miniSOG tagged clones are described in the schematic. The desired band sequenced is highlighted in the box. Lanes G41, G46, G55, C17 is gDNA from hESC clones hESC^{G41}, hESC^{G46}, hESC^{G55} and hESC^{C17}, lane WT; gDNA from wild type hESCs, WC; Water control, M; 1kb ladder.

Similarly, integration of the right homology arm for the clones was also confirmed with PCR using specific primers (P90-P22) (Figure 32).

Figure 32: PCR amplification of the Right Homology arm (RHA) integration in hESC clones. Primer pair used for the PCR is described in the schematic. The desired band sequenced is highlighted. Lanes G41, G46, G55, C17 is gDNA from hESC clones hESC^{G41}, hESC^{G46}, hESC^{G55} and hESC^{C17} respectively, lane WT; gDNA from wild type hESCs, M; 1kb ladder.

The PCR products obtained from both left and right homology arms were sequenced to validate the integration.

3.1.12 Screening of potential bi-allelic gene targeting event in hESC clones

In order to check if the clones had mono/bi-allelic targeting, PCR analysis was done using the primers from the genomic region (P40 and P41) as described in Section 3.1.6, Figure 22. PCR analysis of targeted hESC clones revealed that all the clones had a single allele successfully targeted (Figure 33). The obtained PCR product was sequenced to verify for indels.

Figure 33: PCR amplification to screen for potential bi-allelic targeting of the selected hESC clones. The desired band sequenced is highlighted. Lanes G41, G46, G55, C17 is genomic DNA obtained from hESC clones hESC^{G41}, hESC^{G46}, hESC^{G55} and hESC^{C17}, lane WT; gDNA from wild type hESCs, M; 1kb ladder/100bp ladder.

The sequencing of the PCR product of the untargeted allele revealed that clone hESC^{G41} and hESC^{C17} had WT sequence whereas clones hESC^{G46} and hESC^{G55} had 3 nucleotides deletion, which prevented the disruption of the protein reading frame (Figure 34).

Figure 34: Sequence alignment of the untargeted allele of clones hESC^{G41}, hESC^{G46}, hESC^{G55} and hESC^{C17} to the WT hESC. The region highlighted in blue indicates the mismatches/ indels.

3.1.13 Southern blot analysis of the targeted hESC clones

To confirm the genomic location and the copy number of the integrated cassette, Southern blot analysis was performed. Genomic DNA from each clone was extracted (Section 2.3; Method 1) and subjected to restriction digestion with EcoRI enzyme. PCR-derived digoxigenin-labelled probe complementary to the *GFP* gene was used to detect the left homology arms integration.

Human ESC-*TRFI*-GFP-FLAG clone hESC^{G53} was included for the Southern blot analysis. Dr. Shree Harsha selected the clone and performed the confocal imaging and PCR verification of the integrated cassette. It was confirmed by PCR and Western blot analysis that hESC^{G53} had bi-allelic targeting (data not shown).

Based on Southern blotting analysis, although all the GFP clones gave the expected 3kb band (Figure 35a, b), clones hESC^{G41}, hESC^{G46} and hESC^{G55} showed a band at 8kb similar to the size of the positive control while hESC^{G39} showed two non-specific bands (Figure 35b). The 8kb band was seen in all the southern blots performed with these clones irrespective of the type of probe or the restriction enzyme used to digest the DNA. It was concluded that this band is a recombination product of the integration of the entire donor template with the plasmid backbone. However, bi-allele targeted clone hESC^{G53} did not show the 8kb band (Figure 35b).

Figure 35: Southern blot analysis of targeted hESC clones with an internal GFP probe.

a) A representation of the *TRF1* genomic locus with integrated plasmid (GFP-FLAG) showing the position of *EcoRI* restriction sites and expected band size for the left junction with a GFP probe in a Southern blot analysis.

b) Human ESC clones were digested with *EcoRI* and southern blot analysis was performed using probe complementary to GFP gene. The desired band is highlighted. Lane PC1; 10^7 copies of plasmid, lane PC2; 10^8 copies of plasmid, lanes G41, G46, G55, G39, G53, C17 is gDNA obtained from clones hESC^{G41}, hESC^{G46}, hESC^{G55}, hESC^{G39}, hESC^{G53} and hESC^{C17}, lane WT; gDNA of wild type hESC, M; 1kb ladder.

Next, a Southern blot was performed with an external probe (Figure 36a). This resulted in two bands; one corresponding to the WT locus from the untargeted allele (3983bp) and the other to the targeted locus (2951bp for GFP clones and 3167bp for mCherry clone) (Figure 36a, b). The results also indicated that clones hESC^{G41}, hESC^{G46} and hESC^{G55} have single copy integration at the desired locus (Figure 36c).

Figure 36: Southern blot analysis of targeted hESC clones with an external probe.

A representation of the *TRF1* genomic locus with integrated plasmid (either GFP-FLAG or mCherry-miniSOG) showing the position of probe, *EcoRI* restriction sites and expected band size for the a) Targeted allele (3kb) and b) untargeted allele (4kb), when external probe is used for Southern blot analysis. The desired bands for both the targeted and untargeted alleles are highlighted in the box.

c) Southern blot analysis of targeted hESC clones digested with *EcoRI* using the external probe. Lane 10⁷; 10⁷ copies of plasmid, lane 10⁸; 10⁸ copies of plasmid, lanes G41, G46, G55, G39, G53, C17 are gDNA isolated from hESC^{G41}, hESC^{G46}, hESC^{G55}, hESC^{G39}, hESC^{G53} and hESC^{C17} respectively, lane WT; gDNA from wild type hESC, M; 1kb ladder.

Since clone hESC^{G53} has bi-allelic targeting, it resulted in a single band at the targeted locus of the *TRF1* gene. Clones hESC^{G39} and hESC^{C17} exhibit unspecific band at ~5kb.

3.1.14 Immunoblotting of the tagged TRF1 protein with anti-FLAG and anti-TRF1 antibody in targeted hESC clones

Immunoblotting with anti-FLAG antibody to detect the fusion GFP-FLAG tagged TRF1 protein resulted in a band in clone hESC^{G41} but not in hESC^{G46} or hESC^{G55}

(Figure 37). However, the PCR, sequencing, confocal images and southern blot confirm the integration in all the three clones.

Figure 37: A Western Blot analysis for fusion TRF1 in targeted hESC clones probed with anti-FLAG antibody. Lane WT; whole cell lysate of wildtype hESCs, lanes G41, G46, G55, C17 is whole cell lysates of the targeted hESC clones hESC^{G41}, hESC^{G46}, hESC^{G55} and hESC^{C17}, lane M; molecular weight-size marker. β-actin was used as loading control.

Fusion TRF1 was not detectable by the immunoblotting with anti-TRF1 antibody, therefore further optimizations are required in terms of the protocol and type of antibody to be used (Figure 38).

Figure 38: A Western Blot analysis for fusion TRF1 in targeted hESC clones probed with anti-TRF1 antibody. Lane WT; whole cell lysate of wild type hESCs, lanes G41, G46, G55, C17 is whole cell lysates of the targeted hESC clones hESC^{G41}, hESC^{G46}, hESC^{G55} and hESC^{C17}, lane M; molecular weight-size marker. β-actin was used as loading control.

Based on the PCR, Southern and western analysis, the GFP-FLAG clone hESC^{G53} and mCherry clone hESC^{C17} were selected for further studies. However, the confocal images of hESC^{C17} showed a heterogeneous population of cells and the digestion of genomic DNA resulted in multiple bands in the Southern blot analysis. In order to achieve more homogenous clonal population of hESC^{C17}, a subcloning was performed.

3.1.15 Subcloning of hESC-TRF1-mCherry-miniSOG clone hESC^{C17}

Figure 39: Schematic diagram of the mCherry-miniSOG donor plasmid construct used in this study.

In order to generate single cell clonal population of TRF1-mCherry-miniSOG (Figure 39) clone hESC^{C17}, serial dilution of the cells was done in a 96 well plate format. Upon plating, the wells with a single cell were selected that were allowed to expand to form colonies. After a few weeks, confocal imaging was performed to select the desired sub-colonies that were homogeneous in mCherry expression.

3.1.16 Imaging and PCR screening of the hESC^{C17} subclones

Among the mCherry fluorescent protein expressing positive subclones, four were chosen namely C17-A1, C17-A3, C17-A7 and C17-B1 and expanded for further characterization (Figure 40).

Figure 40: Confocal images of hESC^{C17} subclones C17-A1, C17-A3, C17-A7 and C17-B1 expressing mCherry-miniSOG.

In order to test the successful integration of the left and right homology arms to the C-terminus of *TRF1* gene in hESCs, PCR analysis with primers P21 and P73 for left junction/ P90 and P04 for right junction was performed. PCR and subsequent sequencing results indicated that all four subclones have successful integration of the homology arms (Figure 41). PCR for clone C17-A1 that shows faint band (Figure 41a) was repeated (Figure 41b) to obtain sufficient quantity of DNA for sequencing.

Figure 41: a) PCR amplification of the Left (LHA) and Right (RHA) Homology arms integration in hESC^{C17} subclones. The desired PCR band sequenced is highlighted. Lanes A1, A3, A7 and B1 is gDNA from hESC^{C17} subclones C17-A1, C17-A3, C17-A7, C17-B1 respectively, lane WT; gDNA of wild type hESCs, WC; water control, M; 1kb ladder.

b) PCR amplification of hESC^{C17} subclone C17-A1 to obtain adequate DNA for sequencing. The DNA extracted from both lanes was sequenced.

In order to verify the sequence of the whole cassette, PCR was done to amplify the mCherry gene (with primers P40-P73), mCherry-miniSOG junction (PRmC-PGK) and whole miniSOG gene (with primers P72-puro 24 and P72-PGK) (Figure 42a). The PCR results along with sequencing data indicated that all four subclones have successful integration of the entire template (Figure 42b, c, d).

Figure 42: PCR amplification for sequence verification of the integration cassette in hESC^{C17} subclones.

a) Graphical representation of the donor plasmid with primer pairs indicating the regions amplified for sequence verification.

b) PCR amplification of mCherry gene using primer pair P40 and P73.

c) PCR amplification of the mCherry-miniSOG junction using primer pair PRmC-PGK.

d) PCR amplification of the miniSOG-puromycin junction using primer pair P72 and Puro 24Rv and miniSOG-PGK junction using primer pair P72 and PgK.

Lanes A1, A3, A7 and B1 is gDNA from hESC^{C17} subclones C17-A1, C17-A3, C17-A7, C17-B1 respectively, lane WT; gDNA from wild type hESCs, WC; water control, M; 1kb ladder.

3.1.17 PCR to screen for potential mono/bi-allelic targeting in the hESC^{C17} subclones

In order to test whether the subclones have a single or bi-allelic targeting, a similar PCR strategy was performed. The PCR results indicated that all the four subclones had only one allele targeted (Figure 43). Sequencing of the PCR product revealed that all the four subclones had a 4-nucleotide deletion (Figure 44a) at the same location. This resulted in a protein frame shift at the untagged allele and 26 amino acid longer protein than WT TRF1 (Figure 44b).

Figure 43: PCR amplification to verify bi-allelic targeting in hESC^{C17} subclones. Lanes A1, A3, A7 and B1 is gDNA from hESC^{C17} sub clones C17-A1, C17-A3, C17-A7, C17-B1 respectively, lane WT; gDNA from wild type hESCs, lane WC; water control, M; 1kb ladder.

a)

```

ATAAATTCAACAACCGGACAAGTGTGTCATGTTAAAAGACAGATGGAGGACCATGAAGAAAC
|||||
ATAAATTCAACAACCGGACAAGTGTGTCATGTTAAAAGACAGATGGAGGACCATGAAGAAAC

TAAAACTGATTTCCT----CAGCGAAGACTGATTGTGTTTGTAAAAGCTTGATGAAAGGA  Untagged allele
|||||
TAAAACTGATTTCCTCAGACAGCGAAGACTGATTGTGTTTGTAAAAGCTTGATGAAAGGA  WT sequence
|||||

CAGTTAAGTATTTTGATCACTGCATTTTGTGTTGAAACTTGTGTCATTGATGTAATTTAAA
|||||
CAGTTAAGTATTTTGATCACTGCATTTTGTGTTGAAACTTGTGTCATTGATGTAATTTAAA

```

b)

```

WT sequence  KNLRSQVRKYGEKNWSKILLHYKFNNRTSVMLKDRWRTMKKLKLISSDSED
untagged allele KNLRSQVRKYGEKNWSKILLHYKFNNRTSVMLKDRWRTMKKLKLISSAKTDCVCKSLMKGQLSILITAFCLKLVSMLM

```

Figure 44: Sequencing outcome of the untargeted allele PCR product in hESC^{C17} subclones.

- Sequencing of the untargeted allele PCR product indicating indels upstream of the stop codon of *TRF1* gene in the subclones as compared to the WT *TRF1* sequence.
- TRF1 protein sequence of the untargeted allele of the clones indicating additional 26 amino acids as a result of frameshift as compared to the WT TRF1 protein sequence.

3.1.18 Immunoblotting to detect fusion TRF1 protein expression in hESC^{C17} subclones

To detect the expression of the fusion protein in hESC^{C17} subclones, immunoblotting with anti-TRF1 antibody was performed. Unfortunately, the antibody could not detect the fusion protein (Figure 45). Since the antibody is not characterized, no conclusive inference could be made from the immunoblotting.

Figure 45: A Western Blot analysis for TRF1 in targeted hESC clones probed with anti-TRF1 antibody. Lanes A1, A3, A7 and B1 is whole cell lysates of the targeted hESC^{C17} subclones C17-A1, C17-A3, C17-A7 and C17-B1, lane G53 is whole cell lysate of the hESC^{G53} used as positive control, lane WT; whole cell lysate of wild type hESCs, M; molecular weight-size marker. Image of Ponceau staining is shown for equal loading (bottom panel).

3.1.19 DNA damage induction by miniSOG

Upon illumination with blue light, miniSOG generates singlet oxygen (SOG), which is a form of reactive oxygen species (ROS) (Orrenius, Gogvadze et al. 2007). ROS is harmful to the cells as it can cause single and double stranded DNA breaks. This phenomenon has been used to create site-specific oxidative stress merely by exposing the tagged cells to blue light (Qi, Garren et al. 2012), which when persists will lead to apoptosis.

C17-A3 cells, was employed to determine the effect of DNA damage at telomeres on cell survival. By exposing the cells to blue light, the miniSOG tagged to the *TRF1* of clone C17-A3 generated singlet oxygen that could potentially damage or kill the cells by causing oxidative stress.

C17-A3 cells were seeded in 96 well plate format with different confluency (three rows of 5000, 10,000 and 15,000 cells respectively). In order to find the optimal intensity of blue light to illuminate the plate, different intensities (low, moderate and high) were tested keeping the exposure time constant between the three rows (30sec, 1min30sec, 2min30sec, 3min, 3min30sec and 4min30sec). As control, three wells for each row were not exposed to the blue light. After exposure for the aforementioned times, plates were kept in the incubator under normal culture conditions and monitored for 5 days. Based on the morphology and viability of the cells, moderate intensity (543 units) was chosen as the optimal range for future experiments.

Each well of C17-A3 and WT hESCs was exposed to 543 units of blue light for different duration of time. Following exposure, the cells were kept in the incubator under for 5 days and the cell viability was measured on the day 6. The results indicated that beyond 30 sec, blue light exposure caused increased cell death in the miniSOG expressing C17-A3 compared to the WT cells (Figure 46).

Figure 46: Cell survival assay performed in hESC miniSOG expressing clone C17-A3 and hESC WT cells after exposure to blue light. Cells were exposed to blue light for duration as indicated. Each of the triplicate samples are shown in Linechart. Mean value of the triplicates is shown in the Barchart.

With the small area of a 96-well plate, there was a possibility of the blue light spreading to the neighbouring wells. Therefore, to protect the surrounding wells and ensure uniform light exposure, we repeated the experiment in a 24 well format. Each well was exposed to 543 units of blue light for different durations of time. Following exposure, cells were incubated for 5 days and the cell viability was measured on day 6. The results indicated that beyond 4 minutes, C17-A3 is more prone to cell death compared to the WT cells (Figure 47).

Figure 47: Cell survival assay performed in hESC miniSOG expressing clone C17-A3 and hESC WT cells of each row for different exposure times to blue light. The bar chart indicates the mean values of all the three rows.

Although the results are partially convincing, this data was statistically insignificant. It was challenging to conduct the experiment in a stringent way mainly due to factors like unequal numbers of cells and clumping of cells at the fringe of the wells. Moreover, heat generated from the exposure to blue light caused condensation at the lid of the plate resulting in a gradual uneven light exposure.

Chapter 2

3.2.1 Long-term knockdown of HMGA2 does not affect Telomere length

In order to test if the expression of HMGA2 has an effect on telomere length, we carried out the long-term knockdown (kd) of HMGA2 in HT1080 cells. To this end, HT1080 C1 and C2 cell lines that enable Doxycycline drug (Dox) induced shRNA kd of HMGA2 was employed (Yu, Lim et al. 2014). HT1080 C1 and C2 are clonal cell lines derived from the same parental population of HT1080 cells. Upon observation, it was found that they are not different and behave the same way. Hence HT1080 C1 was used for some experiments whereas HT1080 C2 was used for others. Here, Dox was added to HT1080 C1 cells along with the media every day and cell pellets were collected after 8, 16 and 28 days. As control, HT1080 WT cells were also treated with Dox in the same manner. The mean length of these telomeres was further analyzed using Terminal Restriction Fragment (TRF) analysis (Section 2.6). Genomic DNA from the harvested cells were extracted and digested with two frequent cutter enzyme mix (*RsaI* and *HinfI*) that cuts everywhere in the genome leaving telomeres intact. The digested DNA was separated by agarose gel electrophoresis and subsequently processed for southern blot with a telomere probe (Figure 48). The telomere signal (seen as smear) was quantified to find the mean telomere length (MTL).

The long-term knockdown of HMGA2 revealed that HMGA2 does not affect the length of telomeres (Figure 48). The quantification of the blot did not reveal significant differences between the Dox treated and untreated HT1080 C1 cells. Interestingly, HT1080 C1 cells derived from a clonal population was about 1kb smaller than HT1080 WT cells emphasizing the variability in mean telomere length among different clonal cell lines (Figure 48).

Figure 48: TRF analysis after long-term knockdown of HMGA2 using dox treated HT1080 C1 cells compared to dox treated HT1080 WT cells. Dox-; absence of doxycycline, Dox+; presence of doxycycline (kd of HMGA2), HT1080; gDNA extracted from WT HT1080 cells with/without Dox treatment, HT1080 C1; gDNA extracted from HT1080 C1 cells with/without Dox treatment.

3.2.2 Role of HMGA2 at telomeres during HU induced replication stress

DNA damaging agents such as chemotherapeutic drugs aggravate replication fork stalling and induce double strand breaks (DSBs) that activate the DNA damage repair pathways. (Holliday and Tarrant 1972, Kung et al. 1990, Di Leonardo et al. 1994, Waldman et al. 1996, Robles and Adami 1998). Forks are stalled for prolonged period of time depending on the severity of the damage and the ability of the DNA repair mechanisms. Prolonged stalling leads to fork collapse into DSBs that can be quantified using gel electrophoresis.

In order to determine the effect on telomeres during induced replication stress, TRF analysis (Section 2.6) was performed in the absence of HMGA2. HT1080 C1 (Dox-) cells and HT1080 C1 HMGA2 kd (Dox+) cells were treated with 100mM hydroxyurea (HU) for 24h.

HU induces replication fork stalling by depleting the dNTP pool synthesized during replication (Slater 1973). Based on the previous experiments, it was optimized that four days of doxycycline treatment is sufficient for HMGA2 knockdown (Yu, Lim et al. 2014); therefore a Western analysis was carried after 4 days of dox treatment. Western blot analysis indicated ~80% of HMGA2 knockdown (Figure 49). As controls, Dox-HU- and Dox+HU- cells were used. The results obtained with TRF analysis indicated that there is no alteration to the MTL before or after the HU treatment irrespective of HMGA2 expression (Figure 50)

Figure 49: An immunoblot showing the down-regulation of HMGA2 in HT1080 C1 and C2 cells via doxycycline inducible shRNA. C1; Whole cell lysate from HT1080 C1 cells with/without Dox treatment, C2; whole cell lysate from HT1080 C2 cells with/without Dox treatment. β-actin was used as loading control.

Figure 50: TRF analysis of the HMGA2 expressing HT1080 C1 (Dox-) and HMGA2 knockdown (Dox+) cells treated with and without HU (HU+/ HU-). Lane 1 to 4 and 5 to 8 are duplicates.

TRF analysis did not reveal much about the role of HMGA2 at telomeres. Hence, a different approach was required to probe into the effect of HMGA2 during replication stress induced at telomeres.

3.2.3 HMGA2 protects HU induced stalled replication forks from collapsing into DSBs at telomeres and subtelomeres

To determine the formation of DSBs at the telomeres, HT1080 C1 cells were treated with 100mM HU for 24h in presence or absence of dox. Undigested gDNA extracted from the cells was separated on agarose gel electrophoresis and Southern blot using commercially available telomere probe (Section 2.6) was carried out. This probe contains DIG labeled single stranded telomeric repeats that hybridizes to the target telomere sequence. The southern blotting of the HT1080 C1 cells treated with HU, showed increased smearing in HU+Dox+ cells compared to HU+Dox- cells, suggesting increased DSBs (most likely as a result of collapsed forks) upon HMGA2 knockdown (Figure 51).

Figure 51: Southern blot analysis with telomere probe of HU treated HT1080 C1 cells in presence or absence of HMGA2 to analyze the DSBs occurring at telomeres. Dox -; HMGA2 expression in absence of doxycycline, Dox+; HMGA2 knockdown in presence of doxycycline, HU-; gDNA isolated from HT1080 C1 cells without HU treatment; HU+; gDNA isolated from HT1080 C1 cells when treated with HU. Lane 1 to 4 and 5 to 8 are duplicates. Left and right panel represent two independent experiments.

Based on the telomere length of HT1080 C1 cells (Figure 48), breaks below 8kb could have occurred within the subtelomeres or telomeres. This experiment was repeated twice with duplicates each time and it was confirmed that HMGA2 protects the stalled fork against collapsing into DSBs at the subtelomeres and telomeres (Figure 51).

3.2.4 HMGA2 prevents collapse of HU induced stalled forks genome-wide in HT1080 C1 and C2 cells

Due to the limitation of gel electrophoresis to separate DNA fragments up to 20kb, we next employed Pulsed-field gel electrophoresis (PFGE) to study the DSBs at whole chromosomal level. HT1080 C1 and C2 cells were treated with Dox for three days to kd HMGA2 followed by treatment with Dox and 100mM HU concurrently on Day 4. Dox untreated cells were maintained for three days by adding fresh media everyday and treated with 100mM HU for 24h on Day 4. The cells were harvested and PFGE was performed as described in Section 2.5 and the gel was stained with EtBr. Our PFGE data showed that HU+Dox+ cells have increased DSBs compared to HU+Dox- in both HT1080 C1 and C2 cells, indicating the role of HMGA2 as a replication fork chaperone during replication stress therein restoring genomic stability (Figure 52). This result is consistent with the previous study (Yu, Lim et al. 2014)

Figure 52: PFGE analysis followed by EtBr staining of HT1080 C1 and C2 cells to measure the DSBs after HU treatment. Doses are as indicated. HMGA2 expressing (Dox-) and HMGA2 knockdown (Dox+) cells treated with and without HU (HU+/ HU-). Lane 1 to 4 are HT1080 C1 cells and 5 to 8 are HT1080 C2 cells.

3.2.5 HMGA2 prevents HU induced stalled forks against collapse genome-wide in NSCLC H1299 and A549 cells

In order to investigate the fork protective effect of HMGA2 across different cell lines, we chose H1299 and A549 cell lines. H1299^{HMGA2-KO} of H1299 cell line was generated using CRISPR-Cas9 by our collaborators at Novartis, La Jolla.

We confirmed the knockout of HMGA2 with PCR using specific primers (HMGA2 Fw and HMGA2 Rv). Since a previous publication reported polysomy in chr12 of H1299 cell lines (Liu, Lee et al. 2012), the gel-purified PCR product was TA-cloned into pGEM®-T Vector and sequenced (Section 2.10). Sequencing of the PCR product identified the indels causing frameshift mutation in the alleles screened from the knockout cell line (Figure 53). Western analysis also confirmed the absence of HMGA2 protein in KO cells (Figure 54).

Figure 53: Sequence alignment of the TA-cloned PCR products of H1299^{HMGA2-KO} cell line. The alignment shows indels immediate downstream of the start codon of the HMGA2 gene in the H1299^{HMGA2-KO} cell line compared to the HMGA2 gene reference sequence from NCBI and H1299^{Parental} cell line used in the study.

Figure 54: An immunoblot for HMGA2 in H1299^{Parental} and H1299^{HMGA2-KO} cell lines. H1299^{Parental}, whole cell lysate from H1299^{Parental} cells, H1299^{HMGA2-KO}, whole cell lysate from H1299^{HMGA2-KO}; M, molecular weight- size marker. β-actin was used as loading control.

DNA fragmentation analysis using PFGE (Section 2.5) was performed by treating the H1299^{Parental} and H1299^{HMGA2-KO} cells with 100mM HU for 24h. The gel analyzed with PFGE was stained with EtBr. Although the differences between the HU treated cells and the untreated controls of H1299^{Parental} and H1299^{HMGA2-KO} cell lines are weak, differences between the untreated control and 100mM HU treated cells are significant. Moreover, H1299^{HMGA2-KO} cells indicate increased DSBs at >1Mbp range compared to H1299^{Parental} cells when treated with 100mM HU, implying the protective role of HMGA2 in preventing the stalled replication forks from collapsing into DSBs (Figure 55).

Figure 55: PFGE analysis and EtBr staining of parental H1299^{Parental} and H1299^{HMGA2-KO} cells to measure the DSBs after HU treatment. Doses are as indicated. Left and right panels indicate duplicate experiments.

A549 is a carcinoma cell line derived from lung epithelial cells. They do not express endogenous HMGA2. Therefore, stable cell lines expressing recombinant HMGA2 were generated previously in our lab (Summer, Li et al. 2009). Three clones 1.3, 1.5 and 1.6 expressing HMGA2 were used for the study. The expression of HMGA2 in A549 cells and all three clones was confirmed by western blot analysis and consistent to the original finding, clone 1.5 expresses low-level of HMGA2 compared to 1.3 and 1.6 (Figure 56) (Summer, Li et al. 2009).

Figure 56: An immunoblot of HMGA2 in A549, 1.3, 1.5 and 1.6 cell lines. β-actin used as an internal control. A549; whole cell lysate from A549 cells, 1.3, 1.5, 1.6; whole cell lysate from HMGA2 expressing A549 clones.

Parental A549 and clones 1.3, 1.5 and 1.6 stably expressing HMGA2 were treated with 100mM for 24h. The harvested cells were prepared for PFGE as described in Section 2.5. The EtBr stained gel of PFGE analysis confirmed that the expression of recombinant HMGA2 decreased the DSBs generated as a result of HU treatment. Interestingly, clone 1.5 that expresses low level of HMGA2 compared to 1.3 and 1.6 showed increased DSBs among the three clones, yet less DSBs compared to WT A549 (Figure 57). This suggests a HMGA2 expression level dependency for the formation of pathological replication fork associated DSBs.

Figure 57: Ethidium bromide staining of PFGE analysis of DSBs of A549 and HMGA2 expressing cell lines treated with different doses of HU. Doses are as indicated. Left and right panels indicate duplicate experiments.

Based on these DNA fragmentation analysis conducted, we concluded that the established hypersensitivity to HU by HT1080 cells lacking HMGA2 (Yu, Lim et al. 2014), extends to H1299 cells. Furthermore, the recombinant HMGA2 expression in

A549 cells reduces the DSBs occurrence in HU treated cells thus confirming the role of HMGA2 in protecting the stalled forks during replication stress.

3.2.6 Differential chemosensitivity of HMGA2 expressing cells against Topoisomerase I poison (SN38) induced DNA damage

Topoisomerases are responsible to relax the supercoils formed in replicating cells. SN38 prevents the topoisomerase mediated relaxation by blocking the religation of the DNA nicked by Topoisomerase I. Blocking Topoisomerase activity within cells, leads to increased torsional stress and accumulation of Top1cc complexes. These complexes prevent fork progression and upon collision with replication machinery collapse into pathological DSBs (explained in detail in Section 1.3.1). A previous *in vitro* study from our laboratory found that the association between HMGA2 and SN38 (explained in detail in Section 1.3.2). The study showed that HMGA2 potentiates the efficacy of SN38 in trapping Topo I and DNA in a covalent complex (Peter, Yu et al. 2016). This finding hinted at a functional link between HMGA2 and human Topoisomerase I.

This led us to analyze the role of HMGA2 in SN38 mediated replication stress and DSB formation in the same malignant cell lines that were used to test HU induced replication stress.

3.2.7 HMGA2 protects H1299 and A549 cells against SN38 induced DNA damage

H1299^{Parental} and H1299^{HMGA2-KO} cells were treated with SN38 for 48h. Similarly, A549 cells and HMGA2 expressing clones 1.3, 1.5 and 1.6 were treated with the indicated doses of SN38 for 48h. Fresh media containing SN38 was added everyday. Following treatment, cells were harvested and processed for PFGE (Section 2.5) for DSB analysis. The results showed that expression of HMGA2 protects the H1299^{Parental} and HMGA2 expressing 1.3, 1.5 and 1.6 cells against SN38 induced DSBs (Figure 58, 59). As indicated in the western blot (Figure 56), clone 1.5 has low HMGA2 compared to 1.3 and 1.6. The fact that 1.5 has increased (but still reduced

compared to WT) amounts of fragments compared to 1.3 and 1.6, validates the role of HMGA2 in preventing the DSBs formation.

Figure 58: EtBr staining of PFGE analysis to analyze genomic fragments of H1299^{Parental} and H1299^{HMGA2-KO} cell lines after SN38 treatment in a dose dependent manner. Left and right panels indicate duplicate experiments.

Figure 59: Ethidium bromide staining of PFGE analysis to analyze DSBs in HMGA2 nil (A549) and HMGA2 expressing (1.3, 1.5 and 1.6) cells treated with SN38. Doses are as indicated. Left and right panel are biological duplicate experiments.

3.2.8 HMGA2 sensitizes HT1080 C1 and C2 cells against SN38 induced DNA damage

We next used HT1080 C1 and C2 cells to test for SN38 sensitivity. Cells (C1 and C2) were treated with dox for 4 days to kd HMGA2, with addition of indicated doses of SN38 in last 48h. As a control, the dox untreated cells (C1 and C2) were maintained for 4 days by changing media everyday and treated with indicated doses of SN38 in last 48h. Following treatment, cells were harvested for PFGE (Section 2.5) to analyze the DSB occurrence. Surprisingly, the HMGA2 knockdown cells had decreased DSBs compared to the cells that express HMGA2 (Figure 60). This is in contrast with our finding based on HU induced replication stress in all the four model systems and SN38 mediated replication stress in H1299 and A549 cells.

Figure 60: Ethidium bromide staining of PFGE analysis of HT1080 C1 and C2 HMGA2 expressing (Dox-) and HMGA2 knockdown (Dox+) cells after SN38 treatment in a dose dependent manner.

3.2.9 Differential role of HMGA2 in SN38 induced chemosensitivity impacts cell survival

Beyond a certain threshold, when the cells are unable to cope with the accumulated DSBs, instead of trying to repair the damage, they undergo apoptosis (Kaina 2003). Hence it is reasonable to expect that the differential role of HMGA2 in SN38 induced chemosensitivity observed in PFGE analysis would also affect the cell viability.

3.2.9.1 HMGA2 prevents apoptosis in SN38 treated A549 cells

Ideally, prevention of DSB formation during replication should prevent cells from undergoing apoptosis. Triggering apoptosis would initiate caspase 3/7 activities within cells, which can be measured as an marker for apoptosis (Stennicke, Jurgensmeier et al. 1998, Slee, Harte et al. 1999).

In order to investigate the protective effect of HMGA2 against apoptosis, A549, 1.3, 1.5 and 1.6 cells plated in triplicates in 96well plate were treated with indicated doses of SN38 for 24h. The caspase activity was measured following the treatment period. The results indicate that expression of HMGA2 prevents cancer cell predisposition to apoptosis (Figure 61). Cell line A549 that does not express endogenous HMGA2 show increased caspase 3/7 activities compared to the clones expressing HMGA2 (Figure 61).

Figure 61: Colorimetric Caspase 3/7 activity assay, performed on HMGA2 nil A549 and HMGA2 expressing 1.3, 1.5 and 1.6 cells when treated with SN38. SN38 doses are indicated in the chart. Values are the mean $\pm$ S.D. of triplicate samples. Each panel of graph represents triplicate experiments.

3.2.9.2 HMGA2 decreases cell viability in SN38 treated HT1080 C1 cells

To determine the correlation between increased DSBs in HMGA2 expressing HT1080 C1 and cell survival, CCK8 cell viability assay was performed. Firstly, HT1080 C1 cells were treated with dox to knockdown HMGA2, followed by SN38 treatment for 48h in conjunction with Dox treatment. Dox untreated cells were maintained simultaneously by changing fresh media everyday and treated with indicated doses of SN38 for 48h. Cell viability was then measured using CCK8 assay. The results indicated increased cell survival in HMGA2 knockdown cells compared to HMGA2 expressing cells (Figure 62). This correlates with the PFGE analysis that increased DSBs observed in HMGA2 expressing cells are indeed driving the cells to apoptosis.

Figure 62: Cell survival assay, performed in HMGA2 expressing (Dox-) and HMGA2 knockdown (Dox+) HT1080 C1 cells treated with indicated doses of SN38. Values are the mean $\pm$ S.D. of triplicate samples. Each panel of graph represents duplicate experiments.

3.2.10 Expression of HMGA2 does not alter the endogenous expression of Topoisomerases

In order to determine a correlation between the varying responses to SN38 with the differential expression of Topoisomerases associated with HMGA2 expression, immunoblotting was performed to verify the endogenous levels of Topoisomerases type 1B, 2A and 2B in the HT1080 C1, C2 and A549 cell lines.

Immunoblotting to detect endogenous Topo1B was performed in HT1080 C1 and C2 cells after HMGA2 knockdown. Similarly, immunoblotting with HMGA2 negative A549 cells and recombinant HMGA2 expressing 1.3, 1.5 and 1.6 cells was carried out. The results indicated no substantial correlation between the expression level of HMGA2 and Topo1B (Figure 63).

Figure 63: Immunoblots of Topo1B in HT1080 C1/C2 and A549 cell line and HMGA2 expressing clones. HT1080 C1 and C2 HMGA2 expressing (Dox-) and HMGA2 knockdown (Dox+) cells (top panel) and A549 cells (bottom panel) using β -actin as an internal control. HT1080 C1 & C2; whole cell lysate from HT1080 C1 and C2 cell lines, A549; whole cell lysate from A549 cells, 1.3, 1.4, 1.6; whole cell lysate from HMGA2 expressing A549 cell clones.

We next evaluated the possible association between HMGA2 and Topo2A expression. Using HT1080 C1/C2 and A549 cells model systems; an immunoblot was performed to check the endogenous level of Topo2A. The results indicated no consistent correlation between HMGA2 and Topo 2A (Figure 64).

Western blot for Topo2A was challenging to perform, because of nonspecific bands and frequent degradation of the protein. Result of HT1080 C1 and C2 shown in figure indicates that there is no direct correlation between HMGA2 and Topo2A after HMGA2 knockdown (Figure 64). Similarly, immunoblot to detect Top2A revealed that A549 and 1.5 express Topo2A below the detection range of the antibody. After several repeats, we concluded that there is no correlation between HMGA2 and Topo2A expression (Figure 64). Supporting our inference, another study has also shown that HMGA2 does not affect the expression levels of Topo2A (Boo, Lin et al. 2005).

Figure 64: Immunoblots of Topo2A in HT1080 C1/C2 and A549 cell line and HMGA2 expressing clones. HT1080 C1 and C2 HMGA2 expressing (Dox-) and HMGA2 knockdown (Dox+) cells (left panel) and A549 cells using (right panel) β -actin as an internal control. HT1080 C1 & C2; whole cell lysate from HT1080 C1 and C2 cell lines, A549; whole cell lysate from A549 cells, 1.3, 1.4, 1.6; whole cell lysate from HMGA2 expressing A549 cell clones.

Next we tested the correlation between Topo2B with HMGA2. Immunoblot to detect Topo2B in HT1080 C1/C2 and A549 cell systems revealed that there is no correlation between the expression levels of Topo2B with HMGA2 (Figure 65).

Figure 65: Immunoblots of Topo2B in HT1080 C1/C2 and A549 cell line and HMGA2 expressing clones. HMGA2 expressing (Dox-) and HMGA2 knockdown (Dox+) cells (left panel) and A549 (right panel) using β -actin as an internal control. HT1080 C1 & C2; whole cell lysate from HT1080 C1 and C2 cell lines, A549; whole cell lysate from A549 cells, 1.3, 1.4, 1.6; whole cell lysate from HMGA2 expressing A549 cell clones.

Taken together, these results indicate that the level of HMGA2 does not affect the expression of the Topoisomerase enzymes. This convinced us to exclude a role of topoisomerase expression levels in the differential outcome of sensitization or protection that was observed in SN38 treated cell lines.

3.2.11 Differential expression of HMGA2 determines the differential chemosensitivity to Topoisomerase I poison (SN38)

An *in vitro* DNA nicking assay previously performed in our laboratory to track the efficiency of SN38 mediated Topo1cc trapping by HMGA2 showed that the role HMGA2 in stabilizing the complexes were highly dependent on the stoichiometry of available HMGA2 to the DNA (Zhao, Peter et al. 2017). In other words, increased complex formation was observed at moderate amount of HMGA2:DNA ratio compared to that of low or high amounts of HMGA2 per DNA molecule. These findings prompted us to examine the expression level of HMGA2 across all the cell lines used in the study.

Hence, Western blot was performed to compare the HMGA2 expression level across all the cell lines (Figure 66). Cervical carcinoma HeLa cells do not express endogenous HMGA2. Recombinant HMGA2 expressing HeLa cell lines named P2, P8 and P19 were previously generated in our lab. It was shown that clonal cell line P2 expresses high HMGA2 compared to P8 and P19 (Summer, Li et al. 2009). PFGE analysis performed in SN38 treated cells indicated that transgenic HMGA2 expressing cell lines were more sensitive to SN38 genotoxic effect compared to HMGA2 negative HeLa (data not shown). Thus for the comparative analysis, whole cell lysates from P2 cells along with H1299^{Parental}, HT1080 C1, C2, and A549 1.3 were chosen.

The comparative western blot to detect expression levels of HMGA2 across all cell lines indicates that H1299^{Parental} and 1.3 (HMGA2 expressing A549 derivative) have high HMGA2 expression compared to HT1080 C1, C2 and P2 (HMGA2 expressing HeLa derivative) cell lines (Figure 66). Remarkably, expression of HMGA2 in H1299^{Parental} and 1.3 cells protects the cells against SN38 induced DSBs whereas it sensitizes HT1080 C1, C2 and P2 cells.

Figure 66: Comparative immunoblot of HMGA2 expression level across all the cell lines used in the study. β -actin was used as the internal control. Whole cell lysates from H1299^{Parental}, HT1080 C1/C2, HMGA2 expressing A549 clone 1.3, HMGA2 expressing HeLa clone P2 and HMGA2 nil A549 cells (negative control) were loaded. Left panel shows the representative blot and right panel shows the quantification of band intensity with ImageJ after normalization to β -actin. (Note that the HMGA2 band of clones 1.3 and P2 run at a higher molecular weight compared to the endogenous H1299^{Parental} and HT1080 C1/C2 as the recombinant HMGA2 is tagged with histidine affinity tag).

Next, to address the possibility of the role of Topo1B for the contrasting SnN38 mediated genotoxic effects observed, an immunoblot to compare the levels of endogenous Topo 1B over all the cell lines was performed. This result enabled us to rule out the possible role of Topo1B, as the expression levels are comparable across all cell lines (Figure 67).

A previous study in our laboratory (Yu, Lim et al. 2014) has shown through a study in *E. coli* that the fork protection function is intrinsic to the HMGA family and that HMGA1 could partially substitute for HMGA2. In order to determine the role of HMGA1 in the differential phenotype observed, HMGA1 levels were compared across different cell lines. The comparative immunoblot to detect HMGA1 expression levels revealed that, the level of HMGA1 is relatively comparable across cell lines.

Moreover, the cells that do not express endogenous HMGA2, such as A549 and HeLa, express HMGA1 (Figure 67). In the context of our model, this could imply that HMGA1 does not play an important role along with HMGA2 in the protective/sensitizing effects observed in SN38 induced DSBs. The protein may still

have a contributing function, or the partial substitution by HMGA1 is insufficient to impact SN38 induced genotoxicity.

Figure 67: Comparative immunoblot of endogenous Topo 1B and HMGA1 expression level across all cell lines used in the study.

a) Immunoblot of Topo 1B (Top Panel), HMGA1 (middle panel) and β -actin used as an internal control (bottom panel). Whole cell lysates from H1299^{Parental}, HT1080 C1 & C2, HMGA2 expressing A549 clone 1.3, HMGA2 expressing HeLa clone P2 and A549 parental cells were loaded. Left and right panels are two independent duplicate experiments.

b) Quantification of the band intensities with ImageJ after normalization to β -actin. TOP1B (left panel) and HMGA1 (right panel).

In conclusion, our results suggest that the observed differential sensitivity to Topoisomerase I poison is a result of the differential expression of endogenous HMGA2 (Figure 66). As observed in immunoblotting, high HMGA2 expressing H1299 and A549 cells show resistance, whereas moderate HMGA2 expressing HT1080 C1, C2 and HeLa cells show sensitivity to SN38-induced replication stress (Figure 66).

Chapter 3

3.3.1 Subtelomeric regions are hypersensitive to replication stress and under the control of HMGA2

Analysis of the collapsed forks (DSBs) by PFGE as a result of induced replication stress shows a certain pattern in the occurrence of breaks. The 30-150kb ranging pronounced fragments detected through PFGE are consistent for all the cell models tested for both the drugs. This hints at certain regions in the genome that are more vulnerable to replication insult. Depending on its expression level, HMGA2 either protects or sensitizes this region against genotoxic stress. Either way the association of HMGA2 to these regions seems to have a huge impact on genome stability. To identify the loci of these unknown regions, next generation sequencing was performed.

H1299^{Parental} and H1299^{HMGA2-KO} cells were treated with 2 μ M SN38 for 48h. Following treatment, the harvested cells were processed for PFGE (Section 2.5). The ethidium bromide staining of the gel to visualize DNA showed that H1299^{HMGA2-KO} cells had increased DSBs compared to H1299^{Parental} cells consistent with our previous results. The SN38 untreated controls, showing the normal level of damage in these cells, in the absence of induced replication stress is shown Section 3.2.7 in Figure 58. The DNA fragment smear of interest ranging from 30-150kb was excised from both H1299^{Parental} and H1299^{HMGA2-KO} samples (Figure 68). DNA extracted from both the samples was sequenced via Illumina Hi Seq 2500 sequencer. As controls, genomic DNA of untreated H1299^{Parental} and H1299^{HMGA2-KO} cells were extracted and sequenced.

Figure 68: Ethidium bromide staining of the PFGE of SN38 treated H1299^{Parental} and H1299^{HMGA2-KO}. The 30-150kb DSB fragments excised for sequencing is as indicated in the box. H1299^{Parental}, gDNA isolated from H1299^{Parental} cells after treatment with 2 μ M SN38, H1299^{HMGA2-KO}, gDNA isolated from H1299^{HMGA2-KO} cell line after treatment with 2 μ M SN38.

The sequencing data obtained were checked for quality and the sequencing adapters were trimmed. After the adapter removal, the quality of the data was checked again. The reads were normalized to the read counts for each sample and aligned to the human genome version GRCh38. In order to find the genomic locus of the regions that are enriched in SN38 induced DSBs, normalization of the SN38 treated to the untreated control samples were done for both H1299^{Parental} and H1299^{HMGA2-KO} respectively. The resulting data was binned at 20kb.

The q-value to determine the actual enriched bins by eliminating false positives was calculated statistically using FDR (False Discovery Rate) tools for both H1299^{Parental} and H1299^{HMGA2-KO} samples. Bins enriched upon SN38 treatment (treated versus untreated) was plotted for both H1299^{Parental} and H1299^{HMGA2-KO} samples (Figure 69). FDR cut off of <0.01 was used for identifying significantly enriched or depleted regions for both the samples. 1% FDR is 1.686 fold enrichment for H1299^{Parental} and 1.465 fold enrichment for the H1299^{HMGA2-KO} cell line (Figure 69) (Contributed by Dr. Roland Ivanyi-Nagi).

Figure 69: Distribution of the enriched bins in SN38 induced DSB fragments of H1299^{Parental} (left panel) and H1299^{HMGA2-KO} (right panel) samples. The significantly enriched or depleted regions are highlighted in red for both the samples.

Next, the coverage plots both H1299^{Parental} and H1299^{HMGA2-KO} for each chromosome was plotted. Based on FDR adjusted enrichment thresholds, certain regions were chosen for further analysis (Figure 70). The plots indicated that the subtelomeres and telomeres were significantly enriched in the DSBs fraction.

Figure 70: Coverage plot (chr17 shown as representative).

In order to find the enrichment distribution pattern at subtelomeres and telomeres, the first and last 100kb of each chromosome, which consists of the telomeres and subtelomeres, were plotted (Figure 71). The plot indicates enrichment at these regions.

Figure 71: Enrichment distribution of the first/last 100kb of H1299^{Parental} (left panel) and H1299^{HMGA2-KO} (right panel) SN38 versus untreated samples.

With the extracted enriched 20kb bins values, a genome wide Circos plot was generated (Figure 72). The Circos heat map indicated that in both H1299^{Parental} and H1299^{HMGA2-KO} sample, most of the chromosomes have either one or both telomeric/subtelomeric regions enriched (black) (circles 1 and 3), while the majority has the centromeric region depleted (red) (circles 2 and 4). The genome wide analysis of the enrichment of H1299^{Parental} over H1299^{HMGA2-KO} showed similar distribution except for two large distinct chromosomal regions in chr9 and chrX, which are depleted in the DSB fraction of H1299^{HMGA2-KO} cells (circles 5 and 6). This difference in chr9 and chrX observed is independent of the copy number variation (circle 7). At this point the reason behind this exception in chr9 and chrX are unknown. More studies would be required to interpret this. The enlarged region in chr1 were plotted with 500bp binning, which clearly identifies the enrichment of the telomeric and subtelomeric region upon SN38 treatment (Figure 72).

Figure 72: Circos plot showing genome-wide distribution of differential enrichment and depletion in the SN38 induced DSB fraction of H1299^{Parental} and H1299^{HMGA2-KO} cell lines. Enlarged within (500bp bin size) represents subtelomeric region of chr1 for both samples. With q value set at <0.01, the enrichment and depletion in the H1299^{Parental} (circle1, 2) and H1299^{HMGA2-KO} (Circle 3, 4) are indicated. The enrichment and depletion of H1299^{Parental} over H1299^{HMGA2-KO} is indicated in circle 5 and 6. Copy number variation in H1299^{HMGA2-KO} cell line is represented in circle 7 with purple implying an increase and green implying reduction.

The Circos plot (Figure 72) was generated with the enriched fragments binned at 20kb. Next, in order to find the precise loci of the DSB start and end, sequence reads of the first and last 100kb of each chromosome was binned at 2000bp for the treated and untreated samples of both H1299^{Parental} and H1299^{HMGA2-KO} cell lines. The coverage plot of the treated versus the untreated conditions for each sample gave a succinct locus for each hit. The coordinates obtained were used to visualize the reads of all the four samples (treated and untreated of H1299^{Parental} and H1299^{HMGA2-KO}) in the UCSC genome browser (Figure 73). There was no correlation observed between

the break/junction points and the orientation of the gene transcription units or the GC content of the DSB point regions. As the extracted fragments range from 30-150kb, peaks that were smaller than 20kb in size were eliminated. To visualize the reads by UCSC browser, 500bp binning was used.

Figure 73: Locus of SN38 induced DSBs at the telomeres and subtelomeres. UCSC tracks (500bp bin) of SN38 treated H1299^{Parental}/H1299^{HMGA2-KO} and untreated H1299^{Parental}/H1299^{HMGA2-KO} cells. The red arrows indicate the DNA breakpoints of the drug induced DSBs at the subtelomeres and telomeres. The GC% heatmap is represented below the UCSC tracks. The blue arrows (bottom panel) highlight genes found at corresponding locus based on NCBI RefSeq database.

3.3.2 *Ex vivo* analysis to validate the hypersensitivity of subtelomeres and telomeres towards replication stress

Next generation sequencing analysis of the 30-150kb DSBs, observed in all cell lines during induced replication stress determined that these fragments arise from the telomeric/subtelomeric region. This was verified using Southern blot analysis with a telomere probe. H1299^{Parental} and H1299^{HMGA2-KO} cells treated with indicated doses of SN38 for 48h, were analyzed in PFGE and subsequently processed for Southern blotting (Section 2.6; detection of subtelomeres) (Figure 74). The southern blot data confirmed the sequencing analysis that indeed the 30-150kb fragments are markedly enriched in telomeric and subtelomeric sequences.

Figure 74: Southern blot with telomere probe of H1299^{Parental} and H1299^{HMGA2-KO} cells with indicated doses of SN38 showing the DSB occurrence at subtelomeric and telomeric regions. H1299^{Parental}; gDNA isolated from H1299^{Parental} cells treated with different amounts of SN38, H1299^{HMGA2-KO}; gDNA isolated from H1299^{HMGA2-KO} cells when treated with different amounts of SN38.

To further confirm if telomeres and subtelomeres are sensitized in HT1080 C2 cells treated with SN38 because of the moderate expression level of HMGA2, a Southern blot with telomere probe was performed (Figure 75). This also indicates that the

enriched DSBs seen as a result of topological stress occur at telomeres and subtelomeres (Figure 75). Notably, in HT1080 C2 cells, the range of the fragments is even smaller than 15kb indicating severe damage at both the subtelomeric and telomeric regions

Figure 75: Southern blot with telomere probe in HT1080 C2 HMGA2 expressing (Dox-) and HMGA2 knock down (Dox+) cells with indicated doses of SN38 showing the DSB occurrence at subtelomeric and telomeric regions.

In conclusion, we have identified that the majority of subtelomeric regions in the human genome are exceptionally sensitive to replication stress. The role of HMGA2 in constraining the positive supercoils accumulated as a result of fork stalling prevents fork regression and subsequent pathological DSB formation. We propose that HMGA2 enhances the efficiency of replication fork progression at such vulnerable sites by counteracting the replication stress-induced topological challenges to ensure faithful replication. Interestingly, our data suggests that this protective effect depends on the endogenous expression level of HMGA2, as indicated by the differential sensitivity towards Topoisomerase I poison SN38.

4. DISCUSSION AND FUTURE PROSPECTS

4.1 Generation of *TRF1* tagged stable cell lines using CRISPR/Cas9 genome editing tool

To study the various aspects of telomere biology, cells stably expressing tagged versions of TRF1 from the endogenous locus were generated. Most studies, till date, have focused on over-expression or conditional gene knockout systems to understand the role of shelterin proteins at telomeres. The generation of stable cell lines expressing endogenous levels of TRF1 would be quintessential especially since over-expression of TRF1 has been implicated in telomere shortening and instability (Munoz, Blanco et al. 2009). The variability of telomere length as well as the dynamics and complexity of replication at the telomeres makes it very challenging to perform a straightforward analysis of telomeres. Therefore, the tagging the endogenous *TRF1* with a reporter gene would be useful to facilitate the visualization of telomere movements and to study the interaction of telomere-telomere or telomere-HMGA under different conditions. For this purpose, CRISPR-Cas system was employed to tag the C-terminus of *TRF1* with GFP-FLAG or mCherry-miniSOG in hESCs and HT1080 cells.

Using HT1080 cells, we had obtained clones that had monoallelic *TRF1* gene targeted with single copy of the GFP-FLAG donor plasmid. Immunoblotting confirmed the expression of endogenous TRF1 and fusion protein with FLAG and GFP. The targeted clones did not show any interference to cell growth or cell cycle progression. With targeting of hESCs with GFP-FLAG, we obtained one clone hESC^{G53} with bi-allele targeting at the *TRF1* locus. This clone has been expanded for future studies. The mCherry-miniSOG hESC^{C17}, which was non-clonal, was sub-cloned to obtain clonal cells and characterized. Unexpectedly, the sub-clones also had single allele targeted.

We observed indels in the untargeted allele of the clones obtained from both HT1080 and hESCs at a few bases upstream of the stop codon of the *TRF1* gene, introducing a

frameshift that resulted in a longer than wild-type TRF1. These mutations that are attributed to NHEJ mediated repair of the DSB created by Cas9 at the targeted locus presumably because of the higher nuclease activity of CRISPR/Cas9 is termed as ‘on-target’ mutations (Howden, McColl et al. 2016). However, the position of these indels affects the short, unstructured C-terminal tail of the TRF1 protein and most likely does not alter the functionality of the protein (Fairall, Chapman et al. 2001). The mCherry-miniSOG hESC^{C17} subclones are now being used for telomere chromatin tomography by our collaborator in San Diego.

Furthermore, TRF1 tagged cell lines can be used to study the localization of HMGA proteins at telomeres during normal replication and upon induced replication stress conditions. This tool could also be of immense use to provide evidence for the expanded role of HMGA proteins for instance, its interaction with shelterin components and its role in telomere fragility. With two cell lines being tagged with same proteins, studies comparing HT1080 cancer cell line and human embryonic cells can be performed which would provide valuable insights to the mechanism of action in cancer cells compared to that of pluripotent cells. Hence this tool would enable better understanding of the connection between two important targets of development and disease such as telomeres and HMGA2.

4.2 Effect of oxidative stress at telomeres (proof of concept for miniSOG tagged hESC-TRF1 stable cell line)

With the aim to study the consequences of oxidative stress at telomeres and its effect on cell survival, we used the *TRF1*-mCherry-miniSOG tagged hESC clone C17-A3 to perform photosensitizer experiment that specifically targets telomeres through TRF1. In this experiment, we expect the blue light to excite the miniSOG flavoprotein to generate reactive oxygen species (ROS) at the telomeres (Shu, Lev-Ram et al. 2011). Since it is already known that ROS is detrimental to cells, exceeding the threshold will result in cell death (Orrenius, Gogvadze et al. 2007). Studies have shown that oxidative stress induces single strand breaks preferentially at the telomeres due to it G (guanine) richness and shortens telomeres (von Zglinicki 2002, Coluzzi, Colamartino

et al. 2014). Studies associating telomere shortening and senescence as a result of oxidative stress have been reported (Correia-Melo, Hewitt et al. 2014). Using clone C17-A3, we exposed the cells to various spans of blue light and subsequently after 5 days of recovery, cell viability was measured using the cell survival assay. The experiment was optimized so as to minimize blue light leaking to the neighboring cells causing unintentional damage. The sublethal duration and intensity of blue light exposure were optimized, which allowed us to measure the viability of the miniSOG expressing cells compared to the wildtype human ES cells after recovery period. However, factors such as cell number; unequal distribution of cells in the well and the condensation droplets at the lid of the plate made it difficult to ascertain the findings of the cellular effect from ROS induced stress. Therefore, more trials and optimizations are required to improve our set up to study of effect of oxidative stress via TRF1 induced ROS on telomeres and the overall cell viability. In future, it would be exciting to study how pluripotent cells cope with the ROS stress and DSBs occurring specifically at the telomeres. These cells can be differentiated to cardiomyocytes or neural cells using retinoic acid to study how ROS affects, for example, the differentiation of cells.

4.3 HMGA2 mediated fork protection of the stalled forks under replication stress

DNA replication fidelity is essential for maintaining genome integrity. It is a key cellular process that prevents disease causing mutations and chromosomal aberrations. At each round of replication, the replication fork faces numerous innate impediments such as repeat sequences, secondary structures, topological stress and transcriptionally active domains. These obstructions cause the replication forks to stall, activating DNA repair pathways and leading to cell cycle arrest. A number of proteins are recruited to rescue the stalled forks and when the forks restart the cell cycle resumes as well. This mechanism is compromised in many diseases and genetic disorders. Previous study from our lab had shown that HMGA2 acts as a fork chaperone by forming a scaffold at the stalled forks. This stabilizes the forks, protecting them against endonucleolytic attack that would otherwise lead to irreversible fork collapse, ultimately resulting in DSBs. This has been shown in four

model systems including hESCs and in the HT1080 human fibrosarcoma cell line (Yu, Lim et al. 2014).

Upregulation of HMGA and maintenance of telomere length are common characteristics of ES cells and most cancer cells (Fedele, Battista et al. 2001, Fusco and Fedele 2007). Many independent studies have indicated the role of HMGA proteins in tumorigenesis, but there are only a few studies associating HMGA with telomeres (Dejardin and Kingston 2009, Li, Lin et al. 2011, Bartocci, Diedrich et al. 2014, Natarajan, Begum et al. 2016). Hence, with the recently discovered novel role of HMGA2 at stalled replication forks, we embarked at understanding the function of HMGA2 in the telomeres at a molecular level.

Firstly, to address the question whether HMGA2 knockdown has an effect on telomere length, we carried out telomere Terminal Restriction Fragment (TRF) analysis. Using doxycycline mediated shRNA knock down system for HMGA2 in HT1080 cells, we showed that long term (48 days) down regulation of HMGA2 does not affect the telomere length. This could possibly be due to the insufficient HMGA2 levels to achieve an effect on telomere length as dox induced knockdown of HMGA2 lead to ~70% reduction in protein levels. The comparison of telomere length of the WT mice compared to HMGA2 KO and HMGA1/2 double KO mice generated (Federico, Forzati et al. 2014) could provide answers for the effect of HMGA2 on telomere length.

Next, we investigated the protective role of HMGA2 at the telomeres after HU induced replication stress. Terminal restriction fragment (TRF) analysis showed that there was no difference in the length of telomeres between HMGA2 expressing and knockdown cells after HU treatment. Analysis of the fork protection function of HMGA2 at telomeres was challenging due to the velocity of replication fork. The speed of a replication fork depends on various factors such as species, cell type, locus of the replicon and the stage in S-phase. The rate of the fork is estimated to be ~1-3 kb/min under normal circumstances and ~0.4kb/min in the presence of replication stress. Based on this, the time required to replicate the whole of HT1080 telomeres is ~6 minutes, assuming there is only one ongoing fork per telomere as in the event of nucleotide pool depletion, new forks do not initiate (Conti, Sacca et al. 2007, Techer, Koundrioukoff et al. 2017). The analysis was also unsuccessful because of the fact

that telomeres are highly heterogeneous in length. Increase in lower molecular weight fragments implies an increase in DSBs resulting from fork collapse. In TRF analysis, this is observed as a shift in mean telomere length. To be able to quantify this, majority of the on-going forks at telomeres should be stalled. Hence, the reason behind our unsuccessful attempts to detect the fork chaperone function of HMGA2 possibly could be due to the lack of sufficient number of stalled forks at telomeres.

Following the study of the role of HMGA2 as a telomere length regulator, we next looked at the HMGA2's protective effect at preventing DSBs at telomeric region. We have demonstrated by southern blot analysis that HMGA2 has a protective effect on the stalled forks at and probably within telomeres. We have shown that after HU induced replication stress, there is an increase in the DSBs in the cells where HMGA2 was knocked down, compared to cells expressing endogenous levels of HMGA2. Since the observed DSB fragments range from 1.5kb to over 20kb, it is expected that the fragments arise from the subtelomeric as well as telomeric regions (HT1080 telomeres are ~2 to 8kb) (Figure 76). With this experiment, we provided the first *ex vivo* evidence for HMGA2's protective effect at telomeric and subtelomeric region.

Figure 76: Schematic diagram representing the DSB fragments (indicated in red) occurring at telomeric (indicated in yellow) and subtelomeric (indicated in dark blue) regions after induced replication stress. Since human telomeres range from 2-20kb and subtelomeres from 20-200kb, DSBs occurring within the telomeres would be detected as smear lower than 20kb while DSBs occurring at subtelomeres would be seen as smear less than 200kb in a southern blot performed with a telomere probe.

Owing to the large size of the genomic fragments originating as a result of induced replication stress, we analyzed the DNA fragmentation using PFGE. HMGA2 had previously been shown to promote genome stability and cell survival in HT1080 fibrosarcoma cells by protecting them against HU induced DSBs. We observed this protective effect of HMGA2 in three other cell lines such as H1299, A549 and HeLa (data not shown) as they exhibited hypersensitivity to HU in the absence of HMGA2. A typical stalled fork looks like a Y shaped structure, which is a substrate for DSB as it is prone to endonucleolytic attacks (Lopes, Cotta-Ramusino et al. 2001, Petermann and Helleday 2010). Although the proposed mechanism behind the fork stabilization role of HMGA2 was derived from its known affinity to structures such as chicken-foot and Holliday junction the precise mechanism was not established (Yu, Lim et al. 2014).

4.4 Mechanism of the protective role of HMGA2 at stalled replication forks

Therefore, we next set out to find the mechanism behind the fork protection function of HMGA2. The initial inkling for the mechanism was from the analysis in yeast where HMGA2 expression in yeast lacking *Mec1*, an ATR ortholog prevented fork regression in the presence of HU. Recent studies conducted using *in vitro* assays and single molecule analysis found that HMGA2 curbs both positive and negative DNA supercoiling by binding to the plectonemes with its AT-hooks. The studies suggested that HMGA2 formed a complex with scDNA and interfered with the relaxation by Topoisomerase I (Zhao, Peter et al. 2017). During replication stress, stalled replication forks accumulate positive supercoils ahead of it that needs to be relaxed by the topoisomerases and HMGA2 effectively constrains the supercoils that have escaped relaxation by topoisomerases which in turn prevents the pathological fork regression. On the other hand, in the absence of HMGA2, the positive supercoil build up would lead to fork regression, which would be rapidly converted to DSBs leading to apoptosis (Figure 77) (Neelsen, Zanini et al. 2013, Neelsen and Lopes 2015).

Figure 77: Proposed mechanism showing the role of HMGA2 in HU induced replication stress.

A) Diagram showing on going replication fork and supercoil formation during replication with Topoisomerase I (indicated in blue) binding to supercoils to initiate relaxation.

During HU induced fork stalling, two outcomes that are likely to occur depending on HMGA2 expression and the series of events leading to them are shown in panel B and C.

(Detailed explanation of the initiation of fork reversal is explained in Section 1.3, Figure 7)

These recently acquired data led us to probe the effect of HMGA2 in the cells undergoing topological stress. Hence we looked at the impact of SN38, an irinotecan derivative on cells that lack HMGA2. Our data revealed that the expression of HMGA2 makes certain cancer cell lines sensitive to SN38 while it triggers resistance in some. While absence of HMGA2 in H1299 and A549 cells increased the sensitivity to SN38 induced DSBs, it provided resistance against SN38 in HT1080 C1 and C2 and HeLa (data not shown) cells. Accumulation of DSBs within cells initiates illegal mutations and chromosomal aberrations, ultimately resulting in apoptosis or cell death (Kaina 2003). Hence, we next evaluated the consequences of the observed differential chemosensitivity on cell survival. In compliance with the PFGE data, cell viability assays performed after SN38 treatment confirms that the differential chemosensitivity

indeed impacts cell fate. Our results indicated that the expression of HMGA2, prevents apoptosis in SN38 treated A549 cells, whereas, it decreases cell viability in SN38 treated HT1080 cells.

Immunoblotting to quantify the endogenous Topoisomerase I and II levels did not indicate a strict correlation with HMGA2 expression level. Next, we compared the expression level of HMGA2 across the cell lines we used in our study. We found that H1299^{Parental} and clone 1.5 of A549 cell line have high HMGA2 expression compared to HT1080 C1/C2 and HeLa. Interestingly high HMGA2 expressing cell lines (H1299^{Parental} and 1.5) are resistant to SN38 induced DSBs while the moderate HMGA2 expressing cell lines (HT1080 C1/C2 and HeLa) show sensitivity to SN38 induced DSBs. Cell survival and apoptotic assays performed correlate with the DNA fragmentation analysis revealing that differential expression level of HMGA2 has contrasting effects on cell fate.

Single molecule analysis has revealed that HMGA2 binds to the supercoiled DNA and suppresses the relaxation by Topoisomerase I (Zhao, Peter et al. 2017). Through AFM imaging of the supercoiled DNA with a range of HMGA2 concentrations, it has been shown that at moderate/low level, HMGA2 binds to the crossings in the supercoiled DNA while at very high concentrations, it completely covers the supercoils compacting the plectonemes (Zhao, Peter et al. 2017). In a parallel study, an *in vitro* experiment performed to determine the topoisomerase-DNA reversible intermediate called as the Top1cc complex formation in the presence of SN38 and various amounts of HMGA2, clearly revealed that HMGA2 increases the efficacy of SN38 in trapping the Topoisomerase I and DNA as Top1cc complexes. This depends on the stoichiometry of HMGA2 molecules available for DNA (Peter, Yu et al. 2016). We hypothesized that at high endogenous HMGA2 levels, where cells are resistant to SN38 treatment, the protein covers the entire supercoil seemingly like a condensed DNA. Since HMGA2 is known to compete with Topoisomerase I (Peter, Yu et al. 2016), at high levels of HMGA2, the relaxation of scDNA is repressed preventing the Topoisomerase I binding to supercoils. As a result, SN38 mediated Top1cc complex formation is reduced. On the other hand, supercoil scrunching by HMGA2 prevents the regression of replication forks thus preventing the DSB formation (Figure 78).

At moderate endogenous HMGA2 levels, binding of HMGA2 to the supercoil-interwinds stabilizes the SN38 mediated Top1cc complex formation. The fact that HMGA2 increases the efficiency of SN38, increases the amount of trapped complexes. The uncontrolled accumulation of these complexes at a dangerous intermediate step where the nicked DNA is trapped, leads to the event of protein-linked DNA break (PDB), which is a form of double strand break. Cancer drugs targeting Topoisomerases are known to induce PDBs as a mode of cytotoxicity (Ashour, Atteya et al. 2015). The PDB/DSB can be a result of either collision of the Top1cc with the replisome causing the complex to lose its reversibility or due to the prevention of fork movement because of the positive supercoil build up ahead of the replisome (Figure 78) (D'Arpa, Beardmore et al. 1990, Zhang, D'Arpa et al. 1990, Nitiss and Wang 1996).

In the scenario where cells lack HMGA2, accumulation of unconstrained supercoils as a result of SN38 treatment, leads to deleterious fork regressions, which are converted into DSBs as a result of endonucleolytic attacks. Beyond a certain threshold when cells are unable to repair the DSBs, apoptosis occurs. It is established that topoisomerase I dependent DSBs occur in a dose dependent manner and at low doses of topoisomerase I poison in cancer cells, replication dependent DSBs are solely responsible for cytotoxicity (Hsiang, Lihou et al. 1989, Ashour, Atteya et al. 2015). The dose required to induce DSB differs based on the cell type (dividing versus non-dividing) and exposure time (Figure 78).

Figure 78: Proposed mechanism showing the role of HMGA2 in SN38 induced replication stress.

A) Diagram showing on going replication fork and supercoil formation during replication with Topoisomerase I (indicated in blue) binding to supercoils to initiate relaxation.

During SN38 mediated replication fork stalling, three outcomes that are likely to occur depending on level of HMGA2 expression and the series of events leading to them are shown in panel B, C and D.

(Detailed explanation of the initiation of fork reversal and process of supercoil relaxation by Topoisomerase I are explained in Section 1.3, Figure 7 and Section 1.3.1, Figure 8 respectively)

The conclusions drawn from a combination of studies involving single molecule analysis, *in vitro* assays and *ex vivo* experiments were verified by dot blot analysis using an antibody that recognizes denatured Top1cc complexes (Ahmed, Ramani et al. unpublished). The results indicated that H1299^{Parental} cells expressing high level of HMGA2 had less Top1cc complexes formed in the presence of SN38 compared to moderate HMGA2 expressing HeLa P2 (data not shown). Interestingly, colon cancer patient-derived xenografts treated with irinotecan (precursor of SN38) revealed that moderate expression of HMGA2 sensitizes the PDX models to irinotecan (Ahmed, Ramani et al. unpublished).

4.5 *In silico* and *ex vivo* analysis to determine the genomic locus of HMGA2-associated regions vulnerable to replication stress

The DNA fragmentation analysis resulted in DSB fragments at a certain range that could be referred to as specific ‘hotspots’, regardless of the type of replication stress induced. The genome fragmentation as a result of HU and SN38 treatment occurred at 30-150kb in all the cell lines used in the study. We set out to investigate the genomic locus of these DSBs that seem to be primarily susceptible to replication stress. Particularly, the role of HMGA2 at these regions seems to be inevitable for cell vitality.

Next-generation sequencing analysis of the isolated 30-150kb SN38 induced DSB fragments from H1299^{Parental} and H1299^{HMGA2-KO} cell lines revealed that these fragments arise from the subtelomeric region. The DSB breakpoints analyzed from the two independent samples occur almost at the same locus in the subtelomeric region besides the occurrence of increased breaks in the absence of HMGA2. With this analysis, we show that HMGA2 is acting at the subtelomeric region to aid the replication of such difficult-to-replicate regions. The depletion of centromeres and enrichment of subtelomeres at majority of chromosomes (indicated by the Circos plot) reinstates the vulnerability of chromosome end regions towards replication/topological stress. To further validate our deep sequencing analysis, Southern blot analysis performed with telomere probe on PFGE of SN38 treated H1299 cells showed enrichment of subtelomeric and telomeric signal at the 30-150kb range. Interestingly, subtelomeric regions are also enriched in HT1080 C2 cells where moderate expression level of HMGA2 potentiates the DSBs occurrence.

Subtelomeres are known to be difficult-to-replicate sites due to its heterochromatic and repetitive nature of sequences, that also make them vulnerable to replication stress including induced torsional stress (Drosopoulos, Kosiyatrakul et al. 2012). Fork arrests are common at subtelomeric regions and are even considered as an adaptive trait to initiate rapid cell cycle arrest in order to protect the other important regions/genes of the chromosomes. Unlike telomeres, subtelomeres are hundreds of kilobases long and are replicated by bi-directional forks yet similar to telomeres challenge replication forks as they are heavily transcribed and have long stretches of

tandem repeats including telomeric (Riethman, Ambrosini et al. 2005, Drosopoulos, Kosiyatrakul et al. 2012). Our data suggests that the same 30-150kb subtelomeric regions from the majority of chromosomes were significantly enriched in the fractions isolated from the two independently SN38 treated samples. This indicates that distinct regions in subtelomeres are vulnerable to topological stress induced DSBs and that the absence of HMGA2 further exposes these specific sites to genotoxic agents leading to increased DSBs.

NHEJ is suppressed at the chromosome ends and the preferred mode of DNA repair is HR, possibly for maintaining genome integrity (Bailey, Meyne et al. 1999, Brault and Autexier 2011). Studies have shown that HR takes place at the termini sequences by scanning up to hundreds of kilobases towards the center of the chromosome (Wang, Smogorzewska et al. 2004). However, in the scenario of topological stress leading to DSBs at subtelomeric and telomeric regions, cells may lose the ability to repair via HR and ultimately resort to NHEJ, causing chromosomal fusions and genome instability.

Furthermore, breaks initiating and terminating at the telomeres are also not reported, as they would be 20kb or less which is smaller than the isolated fragments. However, southern blot analysis provides compelling evidence for the presence of telomeric regions in the fragmented DNA. In some cases, one of the break points of the DSBs are at the subtelomeric region while the other break point remains elusive due to the inability to sequence telomeric repeats. Sequence information of some chromosomes is not entirely available and the telomeric tracts are often missing from the available databases. For this reason, if they occur, we were unable to determine the breakpoints occurring within telomeres.

With our analysis, we show that HMGA2 is functionally closely associated with the subtelomeric and telomeric regions during replication and regardless of the nature of stress. The association of HMGA2 to these fragile sites determines the cell fate and most importantly, indirectly affects the cellular response to replication-targeting drugs. In future, it would be interesting to study the overlap between topologically associated domains and the breakpoints our analysis revealed, as this would provide detailed information on the role of HMGA2 at these domains under topological stress.

The function of HMGA2 at the subtelomeres along with its role in constraining supercoils will open new avenues for a better understanding of HMGA2's implications in development and disease.

5. CONCLUSION

- 1) HT1080 fibrosarcoma cells and hESCs stably expressing GFP-FLAG and hESCs stably expressing mCherry-miniSOG at the C-terminus of *TRF1* gene were successfully generated and characterized. The miniSOG tagged cell line was used to induce oxidative stress at the telomeres and to study its effect on cell viability.
- 2) In another study using HT1080 C1^{shHMGA2} cells, it was shown that cells lacking HMGA2 exhibit increased HU-induced DSBs at the telomeres compared to the cells expressing HMGA2. This further led to the finding of the novel fork chaperone function of HMGA2 at telomeres after replication stress and was validated in three different cancer cell lines including recombinant-HMGA2 expressing A549 cell line. Moreover, it was shown that the degree of fork protection correlates with the expression level of HMGA2.
- 3) It was demonstrated that HMGA2 expressing cells exhibit differential sensitivity towards topological stress induced by Topoisomerase I poison SN38 in different cell lines. While expression of HMGA2 protects H1299 and A549 cells against SN38, it sensitizes HT1080 C1 and C2. The contrasting sensitivity observed, attributes to the differential expression level of endogenous HMGA2 in cell lines.
- 4) The *ex vivo* data shows that high level of HMGA2 expression promotes cell survival by constraining the positive supercoils formed ahead of the replication fork thus preventing the pathological fork regression. On the other hand, moderate HMGA2 expression levels stabilize the SN38-mediated Top1cc formation resulting in increased DSBs. In the absence of HMGA2, unconstrained supercoiling leads to accumulation of fork regression and subsequent DNA DSBs. This has also been supported by *in vitro* data and the single molecule analysis.

- 5) Lastly, it was also found that subtelomeric regions are hypersensitive to topological and replication stress and demonstrated that the protective role of HMGA2 extends to subtelomeric and telomeric region. I have shown that HMGA2 regulates the subtelomeric fork progression where it prevents frequent fork collapse through supercoil constraining.

6. REFERENCES

1. Abe, N., T. Watanabe, M. Sugiyama, H. Uchimura, G. Chiappetta, A. Fusco and Y. Atomi (1999). "Determination of high mobility group I(Y) expression level in colorectal neoplasias: a potential diagnostic marker." *Cancer Res* 59(6): 1169-1174.
2. Adair, J. E., Y. Kwon, G. A. Dement, M. J. Smerdon and R. Reeves (2005). "Inhibition of nucleotide excision repair by high mobility group protein HMGA1." *J Biol Chem* 280(37): 32184-32192.
3. Aguilera, A. and B. Gomez-Gonzalez (2008). "Genome instability: a mechanistic view of its causes and consequences." *Nat Rev Genet* 9(3): 204-217.
4. Ancelin, K., C. Brun and E. Gilson (1998). "Role of the telomeric DNA-binding protein TRF2 in the stability of human chromosome ends." *Bioessays* 20(11): 879-883.
5. Arnoult, N., C. Schluth-Bolard, A. Letessier, I. Drascovic, R. Bouarich-Bourimi, J. Campisi, S. H. Kim, A. Boussouar, A. Ottaviani, F. Magdinier, E. Gilson and A. Londono-Vallejo (2010). "Replication timing of human telomeres is chromosome arm-specific, influenced by subtelomeric structures and connected to nuclear localization." *PLoS Genet* 6(4): e1000920.
6. Ashour, M. E., R. Atteya and S. F. El-Khamisy (2015). "Topoisomerase-mediated chromosomal break repair: an emerging player in many games." *Nat Rev Cancer* 15(3): 137-151.
7. Austin, C. A. and L. M. Fisher (1990). "Isolation and characterization of a human cDNA clone encoding a novel DNA topoisomerase II homologue from HeLa cells." *FEBS Lett* 266(1-2): 115-117.
8. Azzalin, C. M. (2012). "UPF1: a leader at the end of chromosomes." *Nucleus* 3(1): 16-21.
9. Azzalin, C. M., P. Reichenbach, L. Khoriantuli, E. Giulotto and J. Lingner (2007). "Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends." *Science* 318(5851): 798-801.
10. Bagga, R., S. Michalowski, R. Sabnis, J. D. Griffith and B. M. Emerson (2000). "HMG I/Y regulates long-range enhancer-dependent transcription on

- DNA and chromatin by changes in DNA topology." *Nucleic Acids Res* 28(13): 2541-2550.
11. Bailey, S. M., J. Meyne, D. J. Chen, A. Kurimasa, G. C. Li, B. E. Lehnert and E. H. Goodwin (1999). "DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes." *Proc Natl Acad Sci U S A* 96(26): 14899-14904.
 12. Baldassarre, G., S. Battista, B. Belletti, S. Thakur, F. Pentimalli, F. Trapasso, M. Fedele, G. Pierantoni, C. M. Croce and A. Fusco (2003). "Negative regulation of BRCA1 gene expression by HMGA1 proteins accounts for the reduced BRCA1 protein levels in sporadic breast carcinoma." *Mol Cell Biol* 23(7): 2225-2238.
 13. Bandaria, J. N., P. Qin, V. Berk, S. Chu and A. Yildiz (2016). "Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin." *Cell* 164(4): 735-746.
 14. Bartocci, C., J. K. Diedrich, I. Ouzounov, J. Li, A. Piunti, D. Pasini, J. R. Yates, 3rd and E. Lazzerini Denchi (2014). "Isolation of chromatin from dysfunctional telomeres reveals an important role for Ring1b in NHEJ-mediated chromosome fusions." *Cell Rep* 7(4): 1320-1332.
 15. Been, M. D. and J. J. Champoux (1984). "Breakage of single-stranded DNA by eukaryotic type 1 topoisomerase occurs only at regions with the potential for base-pairing." *J Mol Biol* 180(3): 515-531.
 16. Benarroch-Popivker, D., S. Pisano, A. Mendez-Bermudez, L. Lototska, P. Kaur, S. Bauwens, N. Djerbi, C. M. Latrick, V. Fraiser, B. Pei, A. Gay, E. Jaune, K. Foucher, J. Cherfils-Vicini, E. Aeby, S. Miron, A. Londono-Vallejo, J. Ye, M. H. Le Du, H. Wang, E. Gilson and M. J. Giraud-Panis (2016). "TRF2-Mediated Control of Telomere DNA Topology as a Mechanism for Chromosome-End Protection." *Mol Cell* 61(2): 274-286.
 17. Bianchi, M. E. and A. Agresti (2005). "HMG proteins: dynamic players in gene regulation and differentiation." *Curr Opin Genet Dev* 15(5): 496-506.
 18. Blackburn, E. H. and J. G. Gall (1978). "A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*." *J Mol Biol* 120(1): 33-53.

19. Blanco, R., P. Munoz, J. M. Flores, P. Klatt and M. A. Blasco (2007).
"Telomerase abrogation dramatically accelerates TRF2-induced epithelial carcinogenesis." *Genes Dev* 21(2): 206-220.
20. Bodnar, A. G., M. Ouellette, M. Frolkis, S. E. Holt, C. P. Chiu, G. B. Morin, C. B. Harley, J. W. Shay, S. Lichtsteiner and W. E. Wright (1998). "Extension of life-span by introduction of telomerase into normal human cells." *Science* 279(5349): 349-352.
21. Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30(15): 2114-2120.
22. Bonnetterre, J. (1995). "[Topoisomerase I inhibitors. Review of phase II trials with irinotecan (CPT-11) and topotecan]." *Bull Cancer* 82(8): 623-628.
23. Boo, L. M., H. H. Lin, V. Chung, B. Zhou, S. G. Louie, M. A. O'Reilly, Y. Yen and D. K. Ann (2005). "High mobility group A2 potentiates genotoxic stress in part through the modulation of basal and DNA damage-dependent phosphatidylinositol 3-kinase-related protein kinase activation." *Cancer Res* 65(15): 6622-6630.
24. Branzei, D. and M. Foiani (2010). "Maintaining genome stability at the replication fork." *Nat Rev Mol Cell Biol* 11(3): 208-219.
25. Brault, M. E. and C. Autexier (2011). "Telomeric recombination induced by dysfunctional telomeres." *Mol Biol Cell* 22(2): 179-188.
26. Brown, P. O. and N. R. Cozzarelli (1981). "Catenation and knotting of duplex DNA by type 1 topoisomerases: a mechanistic parallel with type 2 topoisomerases." *Proc Natl Acad Sci U S A* 78(2): 843-847.
27. Budzowska, M. and R. Kanaar (2009). "Mechanisms of dealing with DNA damage-induced replication problems." *Cell Biochem Biophys* 53(1): 17-31.
28. Burla, R., M. La Torre and I. Saggio (2016). "Mammalian telomeres and their partnership with lamins." *Nucleus* 7(2): 187-202.
29. Cai, J., G. Shen, S. Liu and Q. Meng (2016). "Downregulation of HMGA2 inhibits cellular proliferation and invasion, improves cellular apoptosis in prostate cancer." *Tumour Biol* 37(1): 699-707.
30. Cesare, A. J. and J. Karlseder (2012). "A three-state model of telomere control over human proliferative boundaries." *Curr Opin Cell Biol* 24(6): 731-738.

31. Chai, W., Q. Du, J. W. Shay and W. E. Wright (2006). "Human telomeres have different overhang sizes at leading versus lagging strands." *Mol Cell* 21(3): 427-435.
32. Champoux, J. J. (2001). "DNA topoisomerases: structure, function, and mechanism." *Annu Rev Biochem* 70: 369-413.
33. Champoux, J. J. and R. Dulbecco (1972). "An activity from mammalian cells that untwists superhelical DNA--a possible swivel for DNA replication (polyoma-ethidium bromide-mouse-embryo cells-dye binding assay)." *Proc Natl Acad Sci U S A* 69(1): 143-146.
34. Chase, M. B., S. B. Haga, W. D. Hankins, D. M. Williams, Z. Bi, J. W. Strovel, C. Obrieht and P. E. Berg (1999). "Binding of HMG-I(Y) elicits structural changes in a silencer of the human beta-globin gene." *Am J Hematol* 60(1): 27-35.
35. Chen, L., X. Zhu, Y. Zou, J. Xing, E. Gilson, Y. Lu and J. Ye (2015). "The topoisomerase II catalytic inhibitor ICRF-193 preferentially targets telomeres that are capped by TRF2." *Am J Physiol Cell Physiol* 308(5): C372-377.
36. Chen, X., J. Lechago, A. Ertan, G. Ergun, R. Verm, M. Bridges, C. Johnson, K. Woods, F. Meriano, M. Chirala and M. Younes (2004). "Expression of the high mobility group proteins HMGI(Y) correlates with malignant progression in Barrett's metaplasia." *Cancer Epidemiol Biomarkers Prev* 13(1): 30-33.
37. Chiappetta, G., V. Avantiagato, R. Visconti, M. Fedele, S. Battista, F. Trapasso, B. M. Merciai, V. Fidanza, V. Giancotti, M. Santoro, A. Simeone and A. Fusco (1996). "High level expression of the HMGI (Y) gene during embryonic development." *Oncogene* 13(11): 2439-2446.
38. Chiappetta, G., G. Botti, M. Monaco, R. Pasquinelli, F. Pentimalli, M. Di Bonito, G. D'Aiuto, M. Fedele, R. Iuliano, E. A. Palmieri, G. M. Pierantoni, V. Giancotti and A. Fusco (2004). "HMGA1 protein overexpression in human breast carcinomas: correlation with ErbB2 expression." *Clin Cancer Res* 10(22): 7637-7644.
39. Chieffi, P., S. Battista, M. Barchi, S. Di Agostino, G. M. Pierantoni, M. Fedele, L. Chiariotti, D. Tramontano and A. Fusco (2002). "HMGA1 and HMGA2 protein expression in mouse spermatogenesis." *Oncogene* 21(22): 3644-3650.

40. Chin, L., S. E. Artandi, Q. Shen, A. Tam, S. L. Lee, G. J. Gottlieb, C. W. Greider and R. A. DePinho (1999). "p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis." *Cell* 97(4): 527-538.
41. Coluzzi, E., M. Colamartino, R. Cozzi, S. Leone, C. Meneghini, N. O'Callaghan and A. Sgura (2014). "Oxidative stress induces persistent telomeric DNA damage responsible for nuclear morphology change in mammalian cells." *PLoS One* 9(10): e110963.
42. Conti, C., B. Sacca, J. Herrick, C. Lalou, Y. Pommier and A. Bensimon (2007). "Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells." *Mol Biol Cell* 18(8): 3059-3067.
43. Correia-Melo, C., G. Hewitt and J. F. Passos (2014). "Telomeres, oxidative stress and inflammatory factors: partners in cellular senescence?" *Longev Healthspan* 3(1): 1.
44. Costantino, L. and D. Koshland (2015). "The Yin and Yang of R-loop biology." *Curr Opin Cell Biol* 34: 39-45.
45. Cotta-Ramusino, C., D. Fachinetti, C. Lucca, Y. Doksani, M. Lopes, J. Sogo and M. Foiani (2005). "Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells." *Mol Cell* 17(1): 153-159.
46. d'Adda di Fagagna, F., P. M. Reaper, L. Clay-Farrace, H. Fiegler, P. Carr, T. Von Zglinicki, G. Saretzki, N. P. Carter and S. P. Jackson (2003). "A DNA damage checkpoint response in telomere-initiated senescence." *Nature* 426(6963): 194-198.
47. D'Arpa, P., C. Beardmore and L. F. Liu (1990). "Involvement of nucleic acid synthesis in cell killing mechanisms of topoisomerase poisons." *Cancer Res* 50(21): 6919-6924.
48. de Lange, T. (2005). "Shelterin: the protein complex that shapes and safeguards human telomeres." *Genes Dev* 19(18): 2100-2110.
49. De Vries, B. B., R. Winter, A. Schinzel and C. van Ravenswaaij-Arts (2003). "Telomeres: a diagnosis at the end of the chromosomes." *J Med Genet* 40(6): 385-398.

50. Dejdardin, J. and R. E. Kingston (2009). "Purification of proteins associated with specific genomic Loci." *Cell* 136(1): 175-186.
51. Denchi, E. L. and T. de Lange (2007). "Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1." *Nature* 448(7157): 1068-1071.
52. Deng, Z., Z. Wang, N. Stong, R. Plasschaert, A. Moczan, H. S. Chen, S. Hu, P. Wikramasinghe, R. V. Davuluri, M. S. Bartolomei, H. Riethman and P. M. Lieberman (2012). "A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection." *EMBO J* 31(21): 4165-4178.
53. Ding, H., M. Schertzer, X. Wu, M. Gertsenstein, S. Selig, M. Kammori, R. Pourvali, S. Poon, I. Vulto, E. Chavez, P. P. Tam, A. Nagy and P. M. Lansdorp (2004). "Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein." *Cell* 117(7): 873-886.
54. Drosopoulos, W. C., S. T. Kosiyatrakul and C. L. Schildkraut (2015). "BLM helicase facilitates telomere replication during leading strand synthesis of telomeres." *J Cell Biol* 210(2): 191-208.
55. Drosopoulos, W. C., S. T. Kosiyatrakul, Z. Yan, S. G. Calderano and C. L. Schildkraut (2012). "Human telomeres replicate using chromosome-specific, rather than universal, replication programs." *J Cell Biol* 197(2): 253-266.
56. Du, W., D. Thanos and T. Maniatis (1993). "Mechanisms of transcriptional synergism between distinct virus-inducible enhancer elements." *Cell* 74(5): 887-898.
57. Edwards, D. N., D. K. Orren and A. Machwe (2014). "Strand exchange of telomeric DNA catalyzed by the Werner syndrome protein (WRN) is specifically stimulated by TRF2." *Nucleic Acids Res* 42(12): 7748-7761.
58. Fairall, L., L. Chapman, H. Moss, T. de Lange and D. Rhodes (2001). "Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2." *Mol Cell* 8(2): 351-361.
59. Falvo, J. V., D. Thanos and T. Maniatis (1995). "Reversal of intrinsic DNA bends in the IFN beta gene enhancer by transcription factors and the architectural protein HMG I(Y)." *Cell* 83(7): 1101-1111.
60. Fedele, M., S. Battista, L. Kenyon, G. Baldassarre, V. Fidanza, A. J. Klein-Szanto, A. F. Parlow, R. Visone, G. M. Pierantoni, E. Outwater, M. Santoro,

- C. M. Croce and A. Fusco (2002). "Overexpression of the HMGA2 gene in transgenic mice leads to the onset of pituitary adenomas." *Oncogene* 21(20): 3190-3198.
61. Fedele, M., S. Battista, G. Manfioletti, C. M. Croce, V. Giancotti and A. Fusco (2001). "Role of the high mobility group A proteins in human lipomas." *Carcinogenesis* 22(10): 1583-1591.
 62. Fedele, M., F. Pentimalli, G. Baldassarre, S. Battista, A. J. Klein-Szanto, L. Kenyon, R. Visone, I. De Martino, A. Ciarmiello, C. Arra, G. Viglietto, C. M. Croce and A. Fusco (2005). "Transgenic mice overexpressing the wild-type form of the HMGA1 gene develop mixed growth hormone/prolactin cell pituitary adenomas and natural killer cell lymphomas." *Oncogene* 24(21): 3427-3435.
 63. Federico, A., F. Forzati, F. Esposito, C. Arra, G. Palma, A. Barbieri, D. Palmieri, M. Fedele, G. M. Pierantoni, I. De Martino and A. Fusco (2014). "Hmga1/Hmga2 double knock-out mice display a "superpygmy" phenotype." *Biol Open* 3(5): 372-378.
 64. Fierro-Fernandez, M., P. Hernandez, D. B. Krimer, A. Stasiak and J. B. Schvartzman (2007). "Topological locking restrains replication fork reversal." *Proc Natl Acad Sci U S A* 104(5): 1500-1505.
 65. Finotti, A., L. Breda, C. W. Lederer, N. Bianchi, C. Zuccato, M. Kleanthous, S. Rivella and R. Gambari (2015). "Recent trends in the gene therapy of beta-thalassemia." *J Blood Med* 6: 69-85.
 66. Flynn, R. L., K. E. Cox, M. Jeitany, H. Wakimoto, A. R. Bryll, N. J. Ganem, F. Bersani, J. R. Pineda, M. L. Suva, C. H. Benes, D. A. Haber, F. D. Boussin and L. Zou (2015). "Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors." *Science* 347(6219): 273-277.
 67. Franco, R., F. Esposito, M. Fedele, G. Liguori, G. M. Pierantoni, G. Botti, D. Tramontano, A. Fusco and P. Chieffi (2008). "Detection of high-mobility group proteins A1 and A2 represents a valid diagnostic marker in post-pubertal testicular germ cell tumours." *J Pathol* 214(1): 58-64.
 68. Fusco, A. and M. Fedele (2007). "Roles of HMGA proteins in cancer." *Nat Rev Cancer* 7(12): 899-910.

69. Germe, T., K. Miller and J. P. Cooper (2009). "A non-canonical function of topoisomerase II in disentangling dysfunctional telomeres." *EMBO J* 28(18): 2803-2811.
70. Gilson, E. and V. Geli (2007). "How telomeres are replicated." *Nat Rev Mol Cell Biol* 8(10): 825-838.
71. Gong, Y. and T. de Lange (2010). "A Shld1-controlled POT1a provides support for repression of ATR signaling at telomeres through RPA exclusion." *Mol Cell* 40(3): 377-387.
72. Greider, C. W. (1999). "Telomeres do D-loop-T-loop." *Cell* 97(4): 419-422.
73. Harley, C. B. (2008). "Telomerase and cancer therapeutics." *Nat Rev Cancer* 8(3): 167-179.
74. Hayflick, L. and P. S. Moorhead (1961). "The serial cultivation of human diploid cell strains." *Exp Cell Res* 25: 585-621.
75. Higgins, N. P., K. Kato and B. Strauss (1976). "A model for replication repair in mammalian cells." *J Mol Biol* 101(3): 417-425.
76. Hill, D. A. and R. Reeves (1997). "Competition between HMG-I(Y), HMG-1 and histone H1 on four-way junction DNA." *Nucleic Acids Res* 25(17): 3523-3531.
77. Horwitz, S. B., C. K. Chang and A. P. Grollman (1971). "Studies on camptothecin. I. Effects of nucleic acid and protein synthesis." *Mol Pharmacol* 7(6): 632-644.
78. Howden, S. E., B. McColl, A. Glaser, J. Vadolas, S. Petrou, M. H. Little, A. G. Elefanty and E. G. Stanley (2016). "A Cas9 Variant for Efficient Generation of Indel-Free Knockin or Gene-Corrected Human Pluripotent Stem Cells." *Stem Cell Reports* 7(3): 508-517.
79. Hsiang, Y. H., M. G. Lihou and L. F. Liu (1989). "Arrest of replication forks by drug-stabilized topoisomerase I-DNA cleavable complexes as a mechanism of cell killing by camptothecin." *Cancer Res* 49(18): 5077-5082.
80. Hsieh, M. H., C. H. Tsai, C. C. Lin, T. K. Li, T. W. Hung, L. T. Chang, L. W. Hsin and S. C. Teng (2015). "Topoisomerase II inhibition suppresses the proliferation of telomerase-negative cancers." *Cell Mol Life Sci* 72(9): 1825-1837.
81. Hultdin, M., E. Gronlund, K. F. Norrback, T. Just, K. Taneja and G. Roos (2001). "Replication timing of human telomeric DNA and other repetitive

- sequences analyzed by fluorescence in situ hybridization and flow cytometry." *Exp Cell Res* 271(2): 223-229.
82. Huth, J. R., C. A. Bewley, M. S. Nissen, J. N. Evans, R. Reeves, A. M. Gronenborn and G. M. Clore (1997). "The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif." *Nat Struct Biol* 4(8): 657-665.
 83. Jones, A. M., R. Mitter, R. Poulson, C. Gillett, A. M. Hanby, I. P. Tomlinson, E. J. Sawyer and C. Phyllodes Tumour (2008). "mRNA expression profiling of phyllodes tumours of the breast: identification of genes important in the development of borderline and malignant phyllodes tumours." *J Pathol* 216(4): 408-417.
 84. Kaina, B. (2003). "DNA damage-triggered apoptosis: critical role of DNA repair, double-strand breaks, cell proliferation and signaling." *Biochem Pharmacol* 66(8): 1547-1554.
 85. Kaul, Z., A. J. Cesare, L. I. Huschtscha, A. A. Neumann and R. R. Reddel (2011). "Five dysfunctional telomeres predict onset of senescence in human cells." *EMBO Rep* 13(1): 52-59.
 86. Kazmierczak, B., S. Wanschura, J. Rosigkeit, K. Meyer-Bolte, K. Uschinsky, R. Haupt, E. F. Schoenmakers, S. Bartnitzke, W. J. Van de Ven and J. Bullerdiek (1995). "Molecular characterization of 12q14-15 rearrangements in three pulmonary chondroid hamartomas." *Cancer Res* 55(12): 2497-2499.
 87. Kettunen, E., S. Anttila, J. K. Seppanen, A. Karjalainen, H. Edgren, I. Lindstrom, R. Salovaara, A. M. Nissen, J. Salo, K. Mattson, J. Hollmen, S. Knuutila and H. Wikman (2004). "Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer." *Cancer Genet Cytogenet* 149(2): 98-106.
 88. Kipling, D. and H. J. Cooke (1990). "Hypervariable ultra-long telomeres in mice." *Nature* 347(6291): 400-402.
 89. Kools, P. F. and W. J. Van de Ven (1996). "Amplification of a rearranged form of the high-mobility group protein gene HMGIC in OsA-CI osteosarcoma cells." *Cancer Genet Cytogenet* 91(1): 1-7.
 90. Krakoff, I. H., N. C. Brown and P. Reichard (1968). "Inhibition of ribonucleoside diphosphate reductase by hydroxyurea." *Cancer Res* 28(8): 1559-1565.

91. Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones and M. A. Marra (2009). "Circos: an information aesthetic for comparative genomics." *Genome Res* 19(9): 1639-1645.
92. Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." *Genome Biol* 10(3): R25.
93. Lehnert, U., V. Reat, M. Weik, G. Zaccai and C. Pfister (1998). "Thermal motions in bacteriorhodopsin at different hydration levels studied by neutron scattering: correlation with kinetics and light-induced conformational changes." *Biophys J* 75(4): 1945-1952.
94. Lenain, C., S. Bauwens, S. Amiard, M. Brunori, M. J. Giraud-Panis and E. Gilson (2006). "The Apollo 5' exonuclease functions together with TRF2 to protect telomeres from DNA repair." *Curr Biol* 16(13): 1303-1310.
95. Li, A. Y., H. H. Lin, C. Y. Kuo, H. M. Shih, C. C. Wang, Y. Yen and D. K. Ann (2011). "High-mobility group A2 protein modulates hTERT transcription to promote tumorigenesis." *Mol Cell Biol* 31(13): 2605-2617.
96. Li, T. K. and L. F. Liu (2001). "Tumor cell death induced by topoisomerase-targeting drugs." *Annu Rev Pharmacol Toxicol* 41: 53-77.
97. Ligon, A. H., S. D. Moore, M. A. Parisi, M. E. Mealiffe, D. J. Harris, H. L. Ferguson, B. J. Quade and C. C. Morton (2005). "Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas." *Am J Hum Genet* 76(2): 340-348.
98. Linardopoulou, E. V., E. M. Williams, Y. Fan, C. Friedman, J. M. Young and B. J. Trask (2005). "Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication." *Nature* 437(7055): 94-100.
99. Lingner, J., T. R. Hughes, A. Shevchenko, M. Mann, V. Lundblad and T. R. Cech (1997). "Reverse transcriptase motifs in the catalytic subunit of telomerase." *Science* 276(5312): 561-567.
100. Liu, J., W. Lee, Z. Jiang, Z. Chen, S. Jhunjhunwala, P. M. Haverty, F. Gnad, Y. Guan, H. N. Gilbert, J. Stinson, C. Klijn, J. Guillory, D. Bhatt, S. Vartanian, K. Walter, J. Chan, T. Holcomb, P. Dijkgraaf, S. Johnson, J. Koeman, J. D. Minna, A. F. Gazdar, H. M. Stern, K. P. Hoefflich, T. D. Wu, J. Settleman, F. J. de Sauvage, R. C. Gentleman, R. M. Neve, D. Stokoe, Z. Modrusan, S. Seshagiri, D. S. Shames and Z. Zhang (2012). "Genome and

- transcriptome sequencing of lung cancers reveal diverse mutational and splicing events." *Genome Res* 22(12): 2315-2327.
101. Liu, Y., H. Liang and X. Jiang (2015). "MiR-1297 promotes apoptosis and inhibits the proliferation and invasion of hepatocellular carcinoma cells by targeting HMGA2." *Int J Mol Med* 36(5): 1345-1352.
 102. Lopes, M., C. Cotta-Ramusino, A. Pelliccioli, G. Liberi, P. Plevani, M. Muzi-Falconi, C. S. Newlon and M. Foiani (2001). "The DNA replication checkpoint response stabilizes stalled replication forks." *Nature* 412(6846): 557-561.
 103. Luderus, M. E., B. van Steensel, L. Chong, O. C. Sibon, F. F. Cremers and T. de Lange (1996). "Structure, subnuclear distribution, and nuclear matrix association of the mammalian telomeric complex." *J Cell Biol* 135(4): 867-881.
 104. Maciejowski, J. and T. de Lange (2017). "Telomeres in cancer: tumour suppression and genome instability." *Nat Rev Mol Cell Biol* 18(3): 175-186.
 105. Maestroni, L., S. Matmati and S. Coulon (2017). "Solving the Telomere Replication Problem." *Genes (Basel)* 8(2).
 106. Mari, F., P. Hermanns, M. L. Giovannucci-Uzielli, F. Galluzzi, D. Scott, B. Lee, A. Renieri, S. Unger, B. Zabel and A. Superti-Furga (2009). "Refinement of the 12q14 microdeletion syndrome: primordial dwarfism and developmental delay with or without osteopoikilosis." *Eur J Hum Genet* 17(9): 1141-1147.
 107. Martinez, P. and M. A. Blasco (2010). "Role of shelterin in cancer and aging." *Aging Cell* 9(5): 653-666.
 108. Martinez, P., M. Thanasoula, A. R. Carlos, G. Gomez-Lopez, A. M. Tejera, S. Schoeftner, O. Dominguez, D. G. Pisano, M. Tarsounas and M. A. Blasco (2010). "Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites." *Nat Cell Biol* 12(8): 768-780.
 109. Martinez, P., M. Thanasoula, P. Munoz, C. Liao, A. Tejera, C. McNees, J. M. Flores, O. Fernandez-Capetillo, M. Tarsounas and M. A. Blasco (2009). "Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice." *Genes Dev* 23(17): 2060-2075.

110. Masciullo, V., G. Baldassarre, F. Pentimalli, M. T. Berlingieri, A. Boccia, G. Chiappetta, J. Palazzo, G. Manfioletti, V. Giancotti, G. Viglietto, G. Scambia and A. Fusco (2003). "HMGA1 protein over-expression is a frequent feature of epithelial ovarian carcinomas." *Carcinogenesis* 24(7): 1191-1198.
111. Meyne, J., R. L. Ratliff and R. K. Moyzis (1989). "Conservation of the human telomere sequence (TTAGGG)_n among vertebrates." *Proc Natl Acad Sci U S A* 86(18): 7049-7053.
112. Miller, K. G., L. F. Liu and P. T. Englund (1981). "A homogeneous type II DNA topoisomerase from HeLa cell nuclei." *J Biol Chem* 256(17): 9334-9339.
113. Miller, K. M., O. Rog and J. P. Cooper (2006). "Semi-conservative DNA replication through telomeres requires Taz1." *Nature* 440(7085): 824-828.
114. Mirkin, E. V. and S. M. Mirkin (2007). "Replication fork stalling at natural impediments." *Microbiol Mol Biol Rev* 71(1): 13-35.
115. Morshedi, A., Z. Ren, J. Li and P. Droge (2013). "Probing into the biological processes influenced by ESC factor and oncoprotein HMGA2 using iPSCs." *Stem Cell Rev* 9(4): 514-522.
116. Munoz, P., R. Blanco, G. de Carcer, S. Schoeftner, R. Benetti, J. M. Flores, M. Malumbres and M. A. Blasco (2009). "TRF1 controls telomere length and mitotic fidelity in epithelial homeostasis." *Mol Cell Biol* 29(6): 1608-1625.
117. Munoz, P., R. Blanco, J. M. Flores and M. A. Blasco (2005). "XPF nuclease-dependent telomere loss and increased DNA damage in mice overexpressing TRF2 result in premature aging and cancer." *Nat Genet* 37(10): 1063-1071.
118. Narita, M., M. Narita, V. Krizhanovsky, S. Nunez, A. Chicas, S. A. Hearn, M. P. Myers and S. W. Lowe (2006). "A novel role for high-mobility group a proteins in cellular senescence and heterochromatin formation." *Cell* 126(3): 503-514.
119. Natarajan, S., F. Begum, J. Gim, L. Wark, D. Henderson, J. R. Davie, S. Hombach-Klonisch and T. Klonisch (2016). "High Mobility Group A2 protects cancer cells against telomere dysfunction." *Oncotarget* 7(11): 12761-12782.
120. Natarajan, S., S. Hombach-Klonisch, P. Droge and T. Klonisch (2013). "HMGA2 inhibits apoptosis through interaction with ATR-CHK1 signaling complex in human cancer cells." *Neoplasia* 15(3): 263-280.

121. Neelsen, K. J. and M. Lopes (2015). "Replication fork reversal in eukaryotes: from dead end to dynamic response." *Nat Rev Mol Cell Biol* 16(4): 207-220.
122. Neelsen, K. J., I. M. Zanini, R. Herrador and M. Lopes (2013). "Oncogenes induce genotoxic stress by mitotic processing of unusual replication intermediates." *J Cell Biol* 200(6): 699-708.
123. Nichols, M. D., K. DeAngelis, J. L. Keck and J. M. Berger (1999). "Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11." *EMBO J* 18(21): 6177-6188.
124. Nikitina, T. and C. L. Woodcock (2004). "Closed chromatin loops at the ends of chromosomes." *J Cell Biol* 166(2): 161-165.
125. Nissen, M. S. and R. Reeves (1995). "Changes in superhelicity are introduced into closed circular DNA by binding of high mobility group protein I/Y." *J Biol Chem* 270(9): 4355-4360.
126. Nitiss, J. L. and J. C. Wang (1996). "Mechanisms of cell killing by drugs that trap covalent complexes between DNA topoisomerases and DNA." *Mol Pharmacol* 50(5): 1095-1102.
127. O'Sullivan, R. J. and J. Karlseder (2010). "Telomeres: protecting chromosomes against genome instability." *Nat Rev Mol Cell Biol* 11(3): 171-181.
128. Ofir, R., A. C. Wong, H. E. McDermid, K. L. Skorecki and S. Selig (1999). "Position effect of human telomeric repeats on replication timing." *Proc Natl Acad Sci U S A* 96(20): 11434-11439.
129. Oganessian, L. and T. M. Bryan (2007). "Physiological relevance of telomeric G-quadruplex formation: a potential drug target." *Bioessays* 29(2): 155-165.
130. Olovnikov, A. M. (1973). "A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon." *J Theor Biol* 41(1): 181-190.
131. Orrenius, S., V. Gogvadze and B. Zhivotovsky (2007). "Mitochondrial oxidative stress: implications for cell death." *Annu Rev Pharmacol Toxicol* 47: 143-183.
132. Peter, S., H. Yu, R. Ivanyi-Nagy and P. Droge (2016). "Cell-based high-throughput compound screening reveals functional interaction between oncofetal HMGA2 and topoisomerase I." *Nucleic Acids Res* 44(22): e162.

133. Petermann, E. and T. Helleday (2010). "Pathways of mammalian replication fork restart." *Nat Rev Mol Cell Biol* 11(10): 683-687.
134. Pfeiffer, V. and J. Lingner (2013). "Replication of telomeres and the regulation of telomerase." *Cold Spring Harb Perspect Biol* 5(5): a010405.
135. Pickett, H. A. and R. R. Reddel (2015). "Molecular mechanisms of activity and derepression of alternative lengthening of telomeres." *Nat Struct Mol Biol* 22(11): 875-880.
136. Piscuoglio, S., I. Zlobec, P. Pallante, R. Sepe, F. Esposito, A. Zimmermann, I. Diamantis, L. Terracciano, A. Fusco and E. Karamitopoulou (2012). "HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma." *Histopathology* 60(3): 397-404.
137. Pommier, Y. (2006). "Topoisomerase I inhibitors: camptothecins and beyond." *Nat Rev Cancer* 6(10): 789-802.
138. Pommier, Y., E. Leo, H. Zhang and C. Marchand (2010). "DNA topoisomerases and their poisoning by anticancer and antibacterial drugs." *Chem Biol* 17(5): 421-433.
139. Porro, A., S. Feuerhahn, J. Delafontaine, H. Riethman, J. Rougemont and J. Lingner (2014). "Functional characterization of the TERRA transcriptome at damaged telomeres." *Nat Commun* 5: 5379.
140. Postow, L., N. J. Crisona, B. J. Peter, C. D. Hardy and N. R. Cozzarelli (2001). "Topological challenges to DNA replication: conformations at the fork." *Proc Natl Acad Sci U S A* 98(15): 8219-8226.
141. Postow, L., C. Ullsperger, R. W. Keller, C. Bustamante, A. V. Vologodskii and N. R. Cozzarelli (2001). "Positive torsional strain causes the formation of a four-way junction at replication forks." *J Biol Chem* 276(4): 2790-2796.
142. Pryde, F. E., H. C. Gorham and E. J. Louis (1997). "Chromosome ends: all the same under their caps." *Curr Opin Genet Dev* 7(6): 822-828.
143. Qi, Y. B., E. J. Garren, X. Shu, R. Y. Tsien and Y. Jin (2012). "Photo-inducible cell ablation in *Caenorhabditis elegans* using the genetically encoded singlet oxygen generating protein miniSOG." *Proc Natl Acad Sci U S A* 109(19): 7499-7504.

144. Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang (2013). "Genome engineering using the CRISPR-Cas9 system." *Nat Protoc* 8(11): 2281-2308.
145. Ray Chaudhuri, A., Y. Hashimoto, R. Herrador, K. J. Neelsen, D. Fachinetti, R. Bermejo, A. Cocito, V. Costanzo and M. Lopes (2012). "Topoisomerase I poisoning results in PARP-mediated replication fork reversal." *Nat Struct Mol Biol* 19(4): 417-423.
146. Reeves, R. (2000). "Structure and function of the HMGI(Y) family of architectural transcription factors." *Environ Health Perspect* 108 Suppl 5: 803-809.
147. Reeves, R. (2001). "Molecular biology of HMGA proteins: hubs of nuclear function." *Gene* 277(1-2): 63-81.
148. Reeves, R. and L. Beckerbauer (2001). "HMGI/Y proteins: flexible regulators of transcription and chromatin structure." *Biochim Biophys Acta* 1519(1-2): 13-29.
149. Rice, S. J., S. C. Lai, L. W. Wood, K. R. Helsley, E. A. Runkle, M. M. Winslow and D. Mu (2013). "MicroRNA-33a mediates the regulation of high mobility group AT-hook 2 gene (HMGA2) by thyroid transcription factor 1 (TTF-1/NKX2-1)." *J Biol Chem* 288(23): 16348-16360.
150. Riethman, H. (2008). "Human subtelomeric copy number variations." *Cytogenet Genome Res* 123(1-4): 244-252.
151. Riethman, H., A. Ambrosini, C. Castaneda, J. Finklestein, X. L. Hu, U. Mudunuri, S. Paul and J. Wei (2004). "Mapping and initial analysis of human subtelomeric sequence assemblies." *Genome Res* 14(1): 18-28.
152. Riethman, H., A. Ambrosini and S. Paul (2005). "Human subtelomere structure and variation." *Chromosome Res* 13(5): 505-515.
153. Rogalla, P., K. Drechsler, G. Frey, Y. Hennig, B. Helmke, U. Bonk and J. Bullerdiek (1996). "HMGI-C expression patterns in human tissues. Implications for the genesis of frequent mesenchymal tumors." *Am J Pathol* 149(3): 775-779.
154. Romer, M. U., S. B. Nygard, I. J. Christensen, S. L. Nielsen, K. V. Nielsen, S. Muller, D. H. Smith, B. Vainer, H. J. Nielsen and N. Brunner (2013). "Topoisomerase 1(TOP1) gene copy number in stage III colorectal cancer patients and its relation to prognosis." *Mol Oncol* 7(1): 101-111.

155. Ruyter-Spira, C. P., A. J. de Groof, J. J. van der Poel, J. Herbergs, J. Masabanda, R. Fries and M. A. Groenen (1998). "The HMGI-C gene is a likely candidate for the autosomal dwarf locus in the chicken." *J Hered* 89(4): 295-300.
156. Ruyter-Spira, C. P., J. Herbergs, E. Limpens, J. A. Marsh, J. J. van der Poel, T. A. Ayoubi and M. A. Groenen (1998). "Nucleotide sequence of the chicken HMGI-C cDNA and expression of the HMGI-C and IGF1 genes in autosomal dwarf chicken embryos." *Biochim Biophys Acta* 1399(1): 83-87.
157. Saharia, A., D. C. Teasley, J. P. Duxin, B. Dao, K. B. Chiappinelli and S. A. Stewart (2010). "FEN1 ensures telomere stability by facilitating replication fork re-initiation." *J Biol Chem* 285(35): 27057-27066.
158. Sami, F., X. Lu, S. Parvathaneni, R. Roy, R. K. Gary and S. Sharma (2015). "RECQ1 interacts with FEN-1 and promotes binding of FEN-1 to telomeric chromatin." *Biochem J* 468(2): 227-244.
159. Sander, J. D. and J. K. Joung (2014). "CRISPR-Cas systems for editing, regulating and targeting genomes." *Nat Biotechnol* 32(4): 347-355.
160. Schneider, R. P., I. Garrobo, M. Foronda, J. A. Palacios, R. M. Marion, I. Flores, S. Ortega and M. A. Blasco (2013). "TRF1 is a stem cell marker and is essential for the generation of induced pluripotent stem cells." *Nat Commun* 4: 1946.
161. Schoeftner, S. and M. A. Blasco (2008). "Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II." *Nat Cell Biol* 10(2): 228-236.
162. Sfeir, A., S. Kabir, M. van Overbeek, G. B. Celli and T. de Lange (2010). "Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal." *Science* 327(5973): 1657-1661.
163. Sfeir, A., S. T. Kosiyatrakul, D. Hockemeyer, S. L. MacRae, J. Karlseder, C. L. Schildkraut and T. de Lange (2009). "Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication." *Cell* 138(1): 90-103.
164. Sgarra, R., E. Maurizio, S. Zammitti, A. Lo Sardo, V. Giancotti and G. Manfioletti (2009). "Macroscopic differences in HMGA oncoproteins post-translational modifications: C-terminal phosphorylation of HMGA2 affects its DNA binding properties." *J Proteome Res* 8(6): 2978-2989.

165. Shi, Z., D. Wu, R. Tang, X. Li, R. Chen, S. Xue, C. Zhang and X. Sun (2016). "Silencing of HMGA2 promotes apoptosis and inhibits migration and invasion of prostate cancer cells." *J Biosci* 41(2): 229-236.
166. Shu, X., V. Lev-Ram, T. J. Deerinck, Y. Qi, E. B. Ramko, M. W. Davidson, Y. Jin, M. H. Ellisman and R. Y. Tsien (2011). "A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms." *PLoS Biol* 9(4): e1001041.
167. Slater, M. L. (1973). "Effect of reversible inhibition of deoxyribonucleic acid synthesis on the yeast cell cycle." *J Bacteriol* 113(1): 263-270.
168. Slee, E. A., M. T. Harte, R. M. Kluck, B. B. Wolf, C. A. Casiano, D. D. Newmeyer, H. G. Wang, J. C. Reed, D. W. Nicholson, E. S. Alnemri, D. R. Green and S. J. Martin (1999). "Ordering the cytochrome c-initiated caspase cascade: hierarchical activation of caspases-2, -3, -6, -7, -8, and -10 in a caspase-9-dependent manner." *J Cell Biol* 144(2): 281-292.
169. Sogo, J. M., M. Lopes and M. Foiani (2002). "Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects." *Science* 297(5581): 599-602.
170. Stadler, G., F. Rahimov, O. D. King, J. C. Chen, J. D. Robin, K. R. Wagner, J. W. Shay, C. P. Emerson, Jr. and W. E. Wright (2013). "Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy." *Nat Struct Mol Biol* 20(6): 671-678.
171. Stennicke, H. R., J. M. Jurgensmeier, H. Shin, Q. Deveraux, B. B. Wolf, X. Yang, Q. Zhou, H. M. Ellerby, L. M. Ellerby, D. Bredesen, D. R. Green, J. C. Reed, C. J. Froelich and G. S. Salvesen (1998). "Pro-caspase-3 is a major physiologic target of caspase-8." *J Biol Chem* 273(42): 27084-27090.
172. Stewart, J. A., M. F. Chaiken, F. Wang and C. M. Price (2012). "Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation." *Mutat Res* 730(1-2): 12-19.
173. Strimmer, K. (2008). "fdrtool: a versatile R package for estimating local and tail area-based false discovery rates." *Bioinformatics* 24(12): 1461-1462.
174. Summer, H., O. Li, Q. Bao, L. Zhan, S. Peter, P. Sathiyathan, D. Henderson, T. Klonisch, S. D. Goodman and P. Droge (2009). "HMGA2 exhibits dRP/AP site cleavage activity and protects cancer cells from DNA-

- damage-induced cytotoxicity during chemotherapy." *Nucleic Acids Res* 37(13): 4371-4384.
175. Takai, H., A. Smogorzewska and T. de Lange (2003). "DNA damage foci at dysfunctional telomeres." *Curr Biol* 13(17): 1549-1556.
 176. Tashiro, S., Y. Nishihara, K. Kugou, K. Ohta and J. Kanoh (2017). "Subtelomeres constitute a safeguard for gene expression and chromosome homeostasis." *Nucleic Acids Res* 45(18): 10333-10349.
 177. Techer, H., S. Koundrioukoff, A. Nicolas and M. Debatisse (2017). "The impact of replication stress on replication dynamics and DNA damage in vertebrate cells." *Nat Rev Genet* 18(9): 535-550.
 178. Ten Hagen, K. G., D. M. Gilbert, H. F. Willard and S. N. Cohen (1990). "Replication timing of DNA sequences associated with human centromeres and telomeres." *Mol Cell Biol* 10(12): 6348-6355.
 179. Trapasso, F., M. Sarti, R. Cesari, S. Yendamuri, K. R. Dumon, R. I. Aqeilan, F. Pentimalli, L. Infante, H. Alder, N. Abe, T. Watanabe, G. Viglietto, C. M. Croce and A. Fusco (2004). "Therapy of human pancreatic carcinoma based on suppression of HMGA1 protein synthesis in preclinical models." *Cancer Gene Ther* 11(9): 633-641.
 180. Tsao, Y. P., A. Russo, G. Nyamuswa, R. Silber and L. F. Liu (1993). "Interaction between replication forks and topoisomerase I-DNA cleavable complexes: studies in a cell-free SV40 DNA replication system." *Cancer Res* 53(24): 5908-5914.
 181. Tutton, S., G. A. Azzam, N. Stong, O. Vladimirova, A. Wiedmer, J. A. Monteith, K. Beishline, Z. Wang, Z. Deng, H. Riethman, S. B. McMahon, M. Murphy and P. M. Lieberman (2016). "Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres." *EMBO J* 35(2): 193-207.
 182. van Overbeek, M. and T. de Lange (2006). "Apollo, an Artemis-related nuclease, interacts with TRF2 and protects human telomeres in S phase." *Curr Biol* 16(13): 1295-1302.
 183. von Zglinicki, T. (2002). "Oxidative stress shortens telomeres." *Trends Biochem Sci* 27(7): 339-344.
 184. Vos, S. M., E. M. Tretter, B. H. Schmidt and J. M. Berger (2011). "All tangled up: how cells direct, manage and exploit topoisomerase function." *Nat Rev Mol Cell Biol* 12(12): 827-841.

185. Wang, J. C. (1971). "Interaction between DNA and an Escherichia coli protein omega." *J Mol Biol* 55(3): 523-533.
186. Wang, J. C. (2002). "Cellular roles of DNA topoisomerases: a molecular perspective." *Nat Rev Mol Cell Biol* 3(6): 430-440.
187. Wang, R. C., A. Smogorzewska and T. de Lange (2004). "Homologous recombination generates T-loop-sized deletions at human telomeres." *Cell* 119(3): 355-368.
188. Wang, X., X. Liu, A. Y. Li, L. Chen, L. Lai, H. H. Lin, S. Hu, L. Yao, J. Peng, S. Loera, L. Xue, B. Zhou, L. Zhou, S. Zheng, P. Chu, S. Zhang, D. K. Ann and Y. Yen (2011). "Overexpression of HMGA2 promotes metastasis and impacts survival of colorectal cancers." *Clin Cancer Res* 17(8): 2570-2580.
189. Wright, W. E. and J. W. Shay (1992). "The two-stage mechanism controlling cellular senescence and immortalization." *Exp Gerontol* 27(4): 383-389.
190. Wright, W. E., V. M. Tesmer, M. L. Liao and J. W. Shay (1999). "Normal human telomeres are not late replicating." *Exp Cell Res* 251(2): 492-499.
191. Ye, J., C. Lenain, S. Bauwens, A. Rizzo, A. Saint-Leger, A. Poulet, D. Benarroch, F. Magdinier, J. Morere, S. Amiard, E. Verhoeven, S. Britton, P. Calsou, B. Salles, A. Bizard, M. Nadal, E. Salvati, L. Sabatier, Y. Wu, A. Biroccio, A. Londono-Vallejo, M. J. Giraud-Panis and E. Gilson (2010). "TRF2 and apollo cooperate with topoisomerase 2alpha to protect human telomeres from replicative damage." *Cell* 142(2): 230-242.
192. Ye, J. Z., J. R. Donigian, M. van Overbeek, D. Loayza, Y. Luo, A. N. Krutchinsky, B. T. Chait and T. de Lange (2004). "TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres." *J Biol Chem* 279(45): 47264-47271.
193. Yu, H., H. H. Lim, N. O. Tjokro, P. Sathiyathan, S. Natarajan, T. W. Chew, T. Klonisch, S. D. Goodman, U. Surana and P. Droge (2014). "Chaperoning HMGA2 protein protects stalled replication forks in stem and cancer cells." *Cell Rep* 6(4): 684-697.
194. Zeman, M. K. and K. A. Cimprich (2014). "Causes and consequences of replication stress." *Nat Cell Biol* 16(1): 2-9.
195. Zhang, H., P. D'Arpa and L. F. Liu (1990). "A model for tumor cell killing by topoisomerase poisons." *Cancer Cells* 2(1): 23-27.

196. Zhang, Z., Q. Wang, F. Chen and J. Liu (2015). "Elevated expression of HMGA1 correlates with the malignant status and prognosis of non-small cell lung cancer." *Tumour Biol* 36(2): 1213-1219.
197. Zhao, X., S. Peter, P. Droge and J. Yan (2017). "Oncofetal HMGA2 effectively curbs unconstrained (+) and (-) DNA supercoiling." *Sci Rep* 7(1): 8440.
198. Zhou, X., K. F. Benson, H. R. Ashar and K. Chada (1995). "Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C." *Nature* 376(6543): 771-774.

RESEARCH ARTICLE

Modulation of chromatin DNA topology by the oncofetal HMGA2 protein controls human Subtelomere stability and determines cancer chemosensitivity.

(Manuscript in preparation)

POSTER

HMGA2 as a key regulator of telomere stability through control of DNA topology.

"Eukaryotic DNA Replication & Genome Maintenance" conference at Cold Spring Harbor Laboratory (2017).