

# **INVESTIGATING THE ROLE OF DNA POLYMERASE THETA IN TELOMERIC STABILITY AND THE ESCAPE FROM CRISIS**

**Alys Naomi Evans**

A thesis submitted for the degree of Doctor of Philosophy (PhD)

March 2020



Division of Cancer and Genetics

School of Medicine

Cardiff University



## ACKNOWLEDGEMENTS

Firstly, I would like to thank Professor Duncan Baird for giving me the opportunity to undertake this project. I have truly loved the experience and am incredibly thankful for your continued support throughout the project. Thank you for creating a lab group with such a supportive and friendly atmosphere, it has been a pleasure to work with everyone.

A very special thank you to Dr Kate Liddiard. I am so grateful to have had you as a supervisor and friend. I simply could not have done this without you. I am so thankful for your unwavering support and kindness over the entire course of my project. I cannot express how much I have appreciated your help and your humour. You are truly inspirational.

A huge thank you to Julia Grimstead and Dr Rhiannon Robinson. You two made sure I was totally comfortable in the lab and I could not have done these experiments without you. Julia, my little welshie, thanks for putting up with me, even when I was a health and safety nightmare and entertaining me with your ridiculous humour. Rhi, you will always be my STELA queen. Hearing your singing and laughter in the lab was an absolute highlight. Please don't ever stop!

A massive thank you to Dr Kez Cleal. You made my task much easier and were always willing to help and always did so with a smile.

Thank you to Helene Geiller, I could not have asked for a better friend to go through this process with! Thanks for your friendship and support and thanks for always listening to my stupid questions and insights into the French rugby team. I wish you all the very best for the future.

Thanks also to Dr Kevin Norris and Dr Greg Ngo, you have both been wonderful to work with. Thank you to Dr Thomas Lewis, Joan Jones, Pat Appleton and Peter Henley. You have all been alright company (rude!).

Thanks to Seb, Grace, Jim and Bill Greenwood. Your friendship has helped me through the hardest of times. Thanks to all my family and friends who have pretended to or have actually taken an interest in this project. I am grateful for your support.

Finally, thank you, Simon. You have always encouraged me to fulfil my potential. Thanks for being by my side throughout this experience.

This project was funded by the National Institutes of Health (NIH), I would therefore like to express my thanks for making this PhD possible.

## SUMMARY

During a telomere crisis, critically shortened telomeres are recognised as DNA double-strand breaks (DSBs) by the DNA damage response apparatus, resulting in telomere fusions which propagate genomic instability. DNA polymerase theta (POLQ) is a crucial component of the alternative non-homologous end-joining (A-NHEJ) repair pathway that is implicated in the formation of intra-chromosomal telomere fusion events. Crisis escape is achieved when telomere lengths stabilise following the re-establishment of telomerase activity, as a consequence of large-scale genomic rearrangements. Thus, POLQ is potentially important in malignant transformation and is over-expressed in diverse cancers and associated with poor clinical outcome.

This study aimed to investigate the impact of POLQ on the capacity of cells to escape a telomere-driven crisis, through exploring the role of POLQ in NHEJ repair mechanisms utilised during crisis and its impact on genome evolution.

The disruption of *POLQ* in two distinct human cell lines resulted in a shorter crisis interval compared with their wild-type (WT) counterparts. A significant increase in inter-chromosomal telomere fusion events was observed for the POLQ-deficient cell lines compared with the WT cell lines. This difference was confirmed for one cell line using high-resolution Illumina paired-end sequencing of crisis-stage fusion amplicons. Conversely, a notable reduction in intra-chromosomal 17p sister chromatid fusions was determined for the POLQ-deficient compared with the WT HCT116 cell line. The differences in inter/intra-chromosomal telomere fusion events was reproduced in experiments employing targeted nucleases to induce telomere specific DSBs. Whole genome sequencing of WT and POLQ-deficient clones before and after crisis revealed the remarkable heterogeneity of POLQ-deficient clones that successfully traversed crisis.

Taken together, these data reveal the essential role that POLQ-dependent A-NHEJ plays in mediating intra-chromosomal telomere fusion and the escape from crisis. These observations may have implications for the use of POLQ as a therapeutic target.

## LIST OF ABBREVIATIONS

ALT	Alternative Lengthening of Telomeres
A-NHEJ	Alternative Non-Homologous End-Joining
ATM	Ataxia Telangiectasia Mutated
ATR	Ataxia Telangiectasia and Rad3 Related
BAM	Binary version of a SAM file
BARD1	BRCA1-Associated Ring Domain protein 1
BER	Base-Excision Repair
BFB	Breakage-Fusion-Bridge
BIR	Break-Induced Replication
BLAST	Basic Local Alignment Search
BLM	Bloom Helicase
bp	Base Pairs
BRCA1	Breast Cancer Type 1 Susceptibility
BRCA2	Breast Cancer Type 2 Susceptibility
BSA	Bovine Serum Albumin
Cas9	CRISPR-associated protein 9
CCKR	Cholecystokinin Receptor
CN	Copy Number
C-NHEJ	Classical Non-Homologous End-Joining
CNV	Copy Number Variation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CST	CTC1-STN1-TEN1
CtIP	CtBP-Interacting Protein
DAPI	4',6-Diamidino-2-phenylindole
dH <sub>2</sub> O	Distilled water
ddH <sub>2</sub> O	Double-Distilled water
DDR	DNA Damage Response
D-loop	Displacement Loop
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNA-PKcs	DNA-Dependent Protein Kinase
DN-hTERT	Dominant Negative Human Telomerase Reverse Transcriptase
dNTP	Deoxyribonucleotide

DSB	Double-strand Break
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic Acid
EOCs	Epithelial Ovarian Cancers
EtBr	Ethidium Bromide
EXO1	Exonuclease 1
FA	Fanconi Anaemia
FANCD2	Fanconi Anaemia group D2 protein
FCS	Foetal Calf Serum
FISH	Fluorescence <i>in situ</i> Hybridisation
gDNA	genomic DNA
GFP	Green Fluorescent Protein
GRCh38	Genome Reference Consortium human build 38
HAP1	Human near-haploid chronic myelogenous leukaemia cell line
HEK293	Human Embryonic Kidney 293 cell line
HCl	Hydrochloric acid
HCT116	Human Colon Carcinoma cell line
HR	Homologous Recombination
H2AX	H2A histone family member X
ICL	DNA Interstrand Cross-Links
IGV	Integrative Genome Browser
IMDM	Iscove's Modified Dulbecco's Medium
ITRs	Interstitial Telomeric Repeat Sequences
kb	Kilobases
kDa	KiloDaltons
Ku	Ku70-Ku80 heterodimer
LIG1	DNA Ligase 1
LIG3	DNA Ligase 3
LIG4	DNA Ligase 4
Mb	Megabases
MH	MicroHomology
MMEJ	Microhomology Mediated End-Joining
MMQ-PCR	Monochrome Multiplex Quantitative Polymerase Chain Reaction
MRN	MRE11-RAD50-NBS1
MW	Molecular Weight
NaCl	Sodium Chloride

NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End-Joining
NHP2	Non-Histone Protein 2
NOP10	Nucleolar Protein 10
ORC	Origin Recognition Complex
PANTHER	Protein ANALYSIS THrough Evolutionary Relationships
PARP	Poly(ADP-Ribose) Polymerase
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PD	Population Doubling
p/dg	Per Diploid Genome
p/hg	Per Haploid Genome
PI3K	Phosphoinositide 3-Kinase
PMSF	Phenylmethylsulfonyl fluoride
POLQ	DNA Polymerase theta
POT1	Protection Of Telomeres 1
Pre-RC	Pre-Replication Complexes
PURO	Puromycin
PVDF	Polyvinylidene Fluoride transfer membrane
QC	Quality Control
Q-FISH	Quantitative Fluorescent <i>in situ</i> Hybridisation
Q-PCR	Quantitative Polymerase Chain Reaction
RAP1	Repressor Activator Protein 1
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RPA	Replication Protein A
RPE1	Human Retinal Epithelial cells
RPMI	Roswell Park Memorial Institute
RT	Reverse Transcription
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
SNV	Single Nucleotide Variants
ss	single-stranded
SSA	Single-Strand Annealing

SSBR	Single-Strand Break Repair
ssDNA	single-stranded DNA
STELA	Single TELOmere Length Analysis
SV	Structural Variant
TAE	Tris-Acetate-EDTA
TALEN	Transcription Activator-Like Effector Nuclease
TBE	Tris-Borate EDTA
TCAB1	Telomerase Cajal Body protein 1
TDP	Timing Decision Point
TEN	Telomerase Essential N-terminal
TERC	Telomerase RNA Component
TERRA	Telomeric Repeat-containing RNA
TERT	Telomerase Reverse Transcriptase
TIN2	TRF1-Interacting Nuclear Factor 2
TNK1	Tankyrase-1
TNK2	Tankyrase-2
<i>TP53</i>	Tumour suppressor <i>P53</i> gene
TPE	Telomere Position Effect
TPE-OLD	Telomere Position Effect Over Long Distance
TPP1	TIN2/PTOP/PIP1
TRAP	Telomerase Repeated Amplification Protocol
TRF	Terminal Restriction Fragment
TRF1	Telomeric Repeat-binding Factor 1
TRF2	Telomeric Repeat-binding Factor 2
T-SCEs	Telomere Sister Chromatid Exchange
TVR	Telomeric Variant Repeat
UV	UltraViolet
VAF	Variant Allele Frequency
WGS	Whole Genome Sequencing
WT	Wild Type
XRCC4	X-ray Repair Cross-Complementing protein 4
53BP1	p53-Binding Protein 1
$\gamma$ H2AX	Gamma-H2A histone family member X (DNA damage marker)

## Contents

ACKNOWLEDGEMENTS.....	II
SUMMARY.....	III
LIST OF ABBREVIATIONS.....	IV

### Chapter 1 Introduction

1.1	TELOMERES & TELOMERASE .....	13
1.1.1	Telomere structure .....	13
1.1.2	Telomere replication .....	17
1.1.3	Telomerase .....	23
1.1.4	Alternative lengthening of telomeres (ALT) .....	24
1.1.5	Telomere-Position Effect (TPE) .....	25
1.1.6	Telomere Repeat-containing RNA (TERRA).....	26
1.2	MEASURING TELOMERE LENGTH AND IDENTIFYING FUSION EVENTS.....	27
1.2.1	Terminal Restriction Fragment (TRF) Analysis .....	27
1.2.2	Quantitative Polymerase Chain Reaction (Q-PCR).....	27
1.2.3	Fluorescence <i>in Situ</i> Hybridisation.....	28
1.2.4	Single Telomere Length Analysis (STELA) .....	29
1.2.5	Detecting telomere fusions .....	30
1.3	TELOMERE DYSFUNCTION AND CANCER.....	32
1.3.1	Telomere-driven crisis .....	33
1.3.2	DSB repair pathways.....	34
1.3.3	DNA repair at telomeres .....	43
1.4	DNA POLYMERASE THETA (POLQ).....	44
1.4.1	Discovery.....	44
1.4.2	POLQ structure.....	45
1.4.3	POLQ function.....	46
1.4.4	POLQ and cancer.....	53
1.5	AIMS OF THE PROJECT .....	58

### Chapter 2 Materials and Methods

2.1	MATERIALS .....	60
2.1.1	Reagents.....	60
2.1.2	Plastic lab equipment.....	60
2.1.3	Equipment/machinery.....	60
2.2	METHODS .....	61
2.2.1	Tissue Culture .....	61

2.3	NUCLEOFECTION ASSAY .....	66
2.3.1	Plasmid extraction .....	66
2.3.2	Transient transfection .....	67
2.4	DNA/PROTEIN EXTRACTION.....	68
2.4.1	DNA Extraction from cells .....	68
2.4.2	DNA Quantification .....	68
2.5	POLYMERASE CHAIN REACTION (PCR) .....	70
2.5.1	Conventional PCR .....	70
2.5.2	STELA Polymerase Chain Reaction (PCR).....	70
2.5.3	Fusion PCR .....	71
2.6	GEL ELECTROPHORESIS .....	73
2.6.1	STELA and Fusion amplicons.....	73
2.6.2	Standard DNA products .....	73
2.7	SOUTHERN BLOTTING AND RADIOLABELLING .....	74
2.7.1	Southern Blotting.....	74
2.7.2	Radiolabelling DNA probe.....	74
2.7.3	Hybridisation.....	75
2.7.4	Visualisation of radiolabelled membranes .....	75
2.7.5	Stripping and re-probing membranes.....	75
2.7.6	Gel Analysis.....	76
2.8	TERMINAL RESTRICTION FRAGMENT (TRF) ASSAY .....	77
2.9	TELOMERASE ASSAY .....	78
2.9.1	Protein Extraction .....	78
2.9.2	Protein Quantification.....	78
2.9.3	Telomerase Repeated Amplification Protocol (TRAP).....	78
2.9.4	Electrophoresis .....	80
2.9.5	Visualisation of TRAP products.....	80
2.10	PROTEIN IDENTIFICATION.....	81
2.10.1	Nuclear extracts.....	81
2.10.2	Western Blotting .....	81
2.10.2.3	Electrophoresis .....	82
2.11	SANGER SEQUENCING.....	86
2.11.1	DNA extraction from agarose gel .....	86
2.11.2	Sanger Sequencing using Eurofins.....	86
2.12	HIGH THROUGHPUT SEQUENCING OF TELOMERE FUSION PRODUCTS... 87	
2.12.1	PCR clean-up .....	87
2.12.2	Illumina HiSeq4000 paired-end High Throughput Sequencing .....	88

2.12.3	Fusion calling using Dysgu .....	88
2.12.4	Manual curation and downstream analysis .....	89
2.13	WHOLE GENOME SEQUENCING .....	91
2.13.1	DNA sequencing and mapping .....	91
2.13.2	Copy Number (CN) detection.....	91
2.13.3	Structural Variant (SV) calling.....	92
2.13.4	Insertion and microhomology identification for WGS.....	93
2.13.5	Calculation of gene content of genome.....	93

**Chapter 3 How does the absence of POLQ impact on the ability of cells to escape a telomere-driven crisis?**

3.1	ABSTRACT .....	94
3.2	INTRODUCTION .....	95
3.3	AIMS OF THE CHAPTER.....	97
3.4	RESULTS .....	98
3.4.1	Sequence verification of <i>POLQ</i> mutations.....	98
3.4.2	Protein verification of <i>POLQ</i> mutations .....	101
3.4.3	HAP1 Cell Line .....	102
3.4.4	HCT116 Cell Line .....	106
3.4.5	RPE1 Cell Line .....	108
3.4.6	Comparison of telomere lengths for WT and <i>POLQ</i> -deficient cell lines .....	110
3.4.7	A Telomere-Driven Crisis in the HAP1 Cell Line .....	112
3.4.8	A Telomere-Driven Crisis in the HCT116 Cell Line .....	128
3.5	DISCUSSION .....	136
3.5.1	Characteristics of <i>POLQ</i> -deficient Cell Lines.....	136
3.5.2	Telomere-driven crisis in HAP1 and HCT116 cell lines .....	138
3.5.3	Conclusions.....	141

**Chapter 4 What is the role of POLQ in the NHEJ pathways utilised for telomere fusion?**

4.1	ABSTRACT .....	142
4.2	INTRODUCTION .....	143
4.3	AIMS OF THE CHAPTER.....	144
4.4	RESULTS .....	145
4.4.1	Calculating the Frequency of Fusion Events.....	147
4.4.2	Fusion analysis of clones that did not reach telomere-driven crisis .....	149
4.4.3	Telomere fusion events in HAP1 <sup>DN-hTERT</sup> crisis clones .....	153
4.4.4	Telomere fusion events in HCT116 <sup>DN-hTERT</sup> clones .....	157
4.4.5	Comparing fusion profiles for HAP1 <sup>DN-hTERT</sup> and HCT116 <sup>DN-hTERT</sup> cell lines passaging telomere-driven crisis .....	163

4.4.6	Transient Transfection of WT and POLQ-deficient cell lines with subtelomere-targeting TALEN pairs .....	165
4.4.7	Transient Transfection of HAP1 cells with TALEN pairs .....	165
4.4.8	Transient Transfection of HCT116 cells with TALEN pairs .....	172
4.4.9	Transient Transfection of RPE1 <sup>hTERT</sup> cells with TALEN pairs .....	187
4.5	DISCUSSION .....	193
4.5.1	POLQ-deficient cell lines associated with increased inter-chromosomal telomere fusion and inhibition of intra-chromosomal telomere fusion events. ....	193
4.5.2	Telomere fusion events during crisis .....	195
4.5.3	Conclusions .....	197

**Chapter 5 How does the loss of POLQ impact on the evolving cancer genome?**

5.1	ABSTRACT .....	199
5.2	INTRODUCTION .....	200
5.3	AIMS OF THE CHAPTER .....	201
5.4	Results.....	202
5.4.1	Sequence analysis of HAP1 <sup>DN-hTERT</sup> fusion amplicons generated during telomere-driven crisis .....	202
5.4.2	Mapping pipeline .....	203
5.4.3	Classification of events identified.....	203
5.4.4	Manual validation of fusion events .....	204
5.4.5	Genomic Fusion Events .....	204
5.4.6	Sequencing Output .....	208
5.4.7	Genomic Events .....	209
5.4.8	16p – 21q fusion events.....	229
5.4.9	‘16p’ and ‘21q’ events .....	234
5.4.10	Overall fusion profiles .....	238
5.4.11	Whole Genome Sequencing Analysis .....	240
5.5	DISCUSSION .....	265
5.5.1	Fusion Amplicon Sequencing Data .....	265
5.5.2	Whole Genome Sequencing of HAP1 <sup>DN-hTERT</sup> clones.....	270
5.5.3	Conclusions .....	272

**Chapter 6 General Discussion**

6.1	SUMMARY .....	274
6.2	GENERAL DISCUSSION.....	276
6.2.1	Telomere lengths in POLQ-deficient cell lines.....	276
6.2.2	Escape from telomere-driven crisis in HAP1 and HCT116 cell lines	279
6.2.3	The role of POLQ in telomere fusion events .....	281

6.3	LIMITATIONS, CONCLUSIONS AND FUTURE DIRECTIONS.....	285
6.3.1	Limitations .....	285
6.3.2	Conclusions.....	286
6.3.3	Future directions.....	287
	<b>Appendix.....</b>	<b>289</b>
	<b>References .....</b>	<b>312</b>

# Chapter 1

## Introduction

### 1.1 TELOMERES & TELOMERASE

The telomere was first described in 1938 by geneticist Hermann Muller. Whilst studying X-ray induced DNA breaks in *Drosophila melanogaster*, he observed that chromosome ends were protected from genetic alterations. Exposing DNA to ionising radiation can cause chromosomal rearrangements, however, Muller noted that these aberrations never involved the ends of chromosomes (Muller 1938). This led to the hypothesis that a specialised structure existed at chromosome ends which had a protective function. A few years later Barbara McClintock made similar observations in *Zea mays* and determined that chromosome ends were able to fuse if this terminal structure was absent (McClintock 1941).

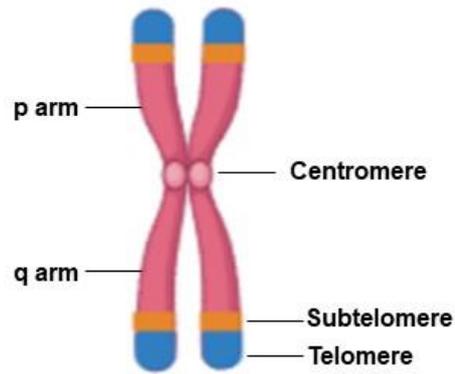
It is now understood that telomeres are the structures found at the ends of linear, eukaryotic chromosomes and are important for maintaining genomic integrity and dictating the replicative lifespan of the cell. Telomeres therefore have a fundamental role in the cell, highlighted by their conservation across eukaryotic organisms. The human telomeric repeat sequence, TTAGGG, was first identified in 1988 and is well conserved amongst vertebrates, (Moyzis et al. 1988) although the telomeric sequence and its length does vary between organisms.

#### 1.1.1 Telomere structure

In humans, telomeres are found at the ends of both the long q-arm and the short p-arm of the chromosome. The hexameric telomere sequence (TTAGGG) is tandemly repeated into arrays, the length of which depends on the individual and the tissue sample analysed (Griffith et al. 1999). Telomere repeats are synthesised by the enzyme telomerase, a holoenzyme with different functional components (Greider and Blackburn 1985). However, in human somatic cells, telomerase is not usually

expressed (Harley et al. 1990) (*Section 1.1.3*). Between each telomere and the chromosome-specific DNA sequence is a sequence referred to as the subtelomeric region which ranges from of 10-300 kilobases (kb) (Riethman et al. 2004) (**Figure 1.1A**). Telomeres are made up regions of repeated double-stranded DNA (dsDNA) sequence. However, the most distal 50–200 base pairs (bp) terminate in a G-rich single-stranded (ss) overhang at the 3' end and a shorter C-rich strand at the 5' end (Greider 1999) (**Figure 1.1B**). The G-rich 3'-overhang (otherwise known as the G-tail) folds back and inserts itself within the preceding double-stranded region of the telomere, forming a T-loop structure (Griffith et al. 1999). This displaces a strand of double-stranded telomeric sequence forming an additional looped structure, known as the displacement loop (D-loop) (Greider 1999) (**Figure 1.1C**).

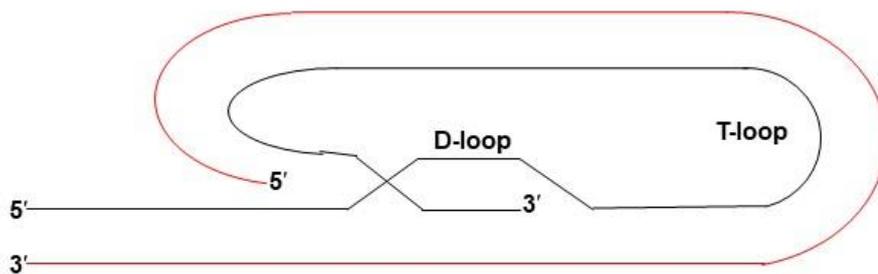
A



B



C



**Figure 1.1:** (A) Telomeres cap the ends of linear chromosomes. The G-rich 3' end forms an overhang (B) which folds back on itself and inserts itself within the preceding double-stranded section of the telomere forming a T-loop structure (C).

### **1.1.1.1 Interstitial telomeric repeat sequences (ITRs)**

Whilst telomeric repeats are found at the chromosome terminus, large blocks of telomere repeats can be found in non-terminal regions in some vertebrate species (Meyne et al. 1990). These interstitial telomeric repeat sequences (ITRs) can be found close to the centromere or at interstitial sites between the centromere and the telomere (Balajee et al. 1994). It is suggested that these non-functional ITRs are the result of chromosome fusion events during evolution or the insertion of telomeric DNA within unstable sites as a consequence of aberrant double-strand break (DSB) repair (Hastie and Allshire 1989).

### **1.1.1.2 Subtelomeric DNA**

Subtelomeric regions are located between the telomere repeat array and the chromosome-specific genomic sequences. Subtelomeres constitute multiple sequence segments that are highly similar to each other but are distinct from the telomere repeat sequence (Tashiro et al. 2017). The subtelomeric regions of eukaryotic chromosomes are highly recombinogenic and this creates extensive structural and sequence variation (Batté et al. 2017). These regions of the genome often contain diverse gene families that can facilitate adaptation to ecologic niches (Mefford and Trask 2002). These include the virulence genes in parasitic eukaryotes, olfactory receptors in mammals and genes involved in disaccharide metabolism in yeast. In humans, there is a high level of sequence homology between the subtelomeric sequences from different chromosome ends. There are therefore groups of related chromosome ends, such as those with homology to 16p or the 21q telomeres (Brown et al. 1990; Letsolo et al. 2010) that include the following chromosome ends:

- 16p family (16p, 1p, 9p, 12p, 15q, XqYq and the 2q14 interstitial locus)

- 21q family (21q, 1q, 2q, 5q, 6q, 6p, 8p, 10q, 13q, 17q, 19p, 19q, 22q and the 2q13 interstitial locus)

### 1.1.1.3 Telomere repeat variants (TVRs)

Degenerate telomeric repeat variants (TVR) occur in the proximal 1-2 kb of the telomere repeat array (Allshire et al. 1989). TVRs display hypervariable interspersed patterns (Baird et al. 1995), and the most common human TVR sequences include TTGGGG, TCAGGG and TGAGGG (Baird et al. 2000). Whilst telomeric DNA preserves genome integrity, TVR sequences do not appear to be associated with this function. Instead, TVR sequences have been associated with unstable telomeres and studies have demonstrated that telomeres lose their protective function when a telomere has shortened to within the TVR region (Capper et al. 2007). Therefore, telomere function is not just defined by its nucleotide sequence but requires several proteins which bind to telomeric DNA. Together, these proteins form the shelterin complex (*Section 1.1.2.1*) which is essential for facilitating the formation and stabilisation of the T- and D-loop structures (Xin et al. 2008).

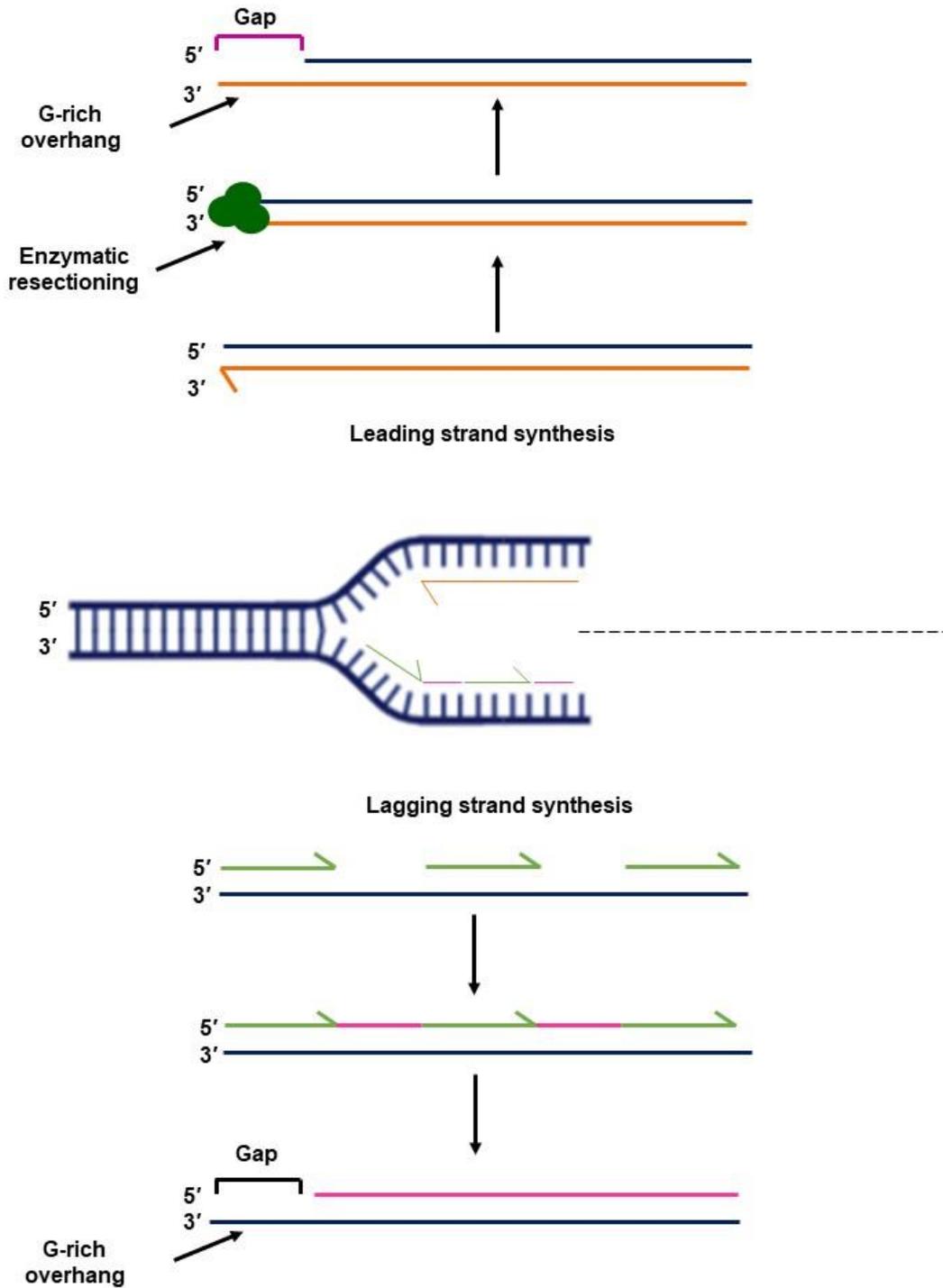
### 1.1.2 Telomere replication

During semi-conservative DNA replication, one of the two newly synthesised DNA strands is made continuously in the 5' to 3' direction and is referred to as the leading strand. As the DNA polymerase involved in replication is unable to replicate DNA in the 3' to 5' direction, discontinuous synthesis is employed to replicate the lagging strand. RNA primers are added to the lagging strand and DNA is synthesised in many small sections called Okazaki fragments. The RNA primers are then degraded, the resulting gaps are filled, and the Okazaki fragments are ligated to form an unbroken strand (O'Sullivan and Karlseder 2010). When the replication fork reaches the end of the chromosome, there is a small portion of DNA that does not get covered by an

Okazaki fragment. This results in a ssDNA overhang that becomes the G-tail (Lundblad 2012). Resectioning of the end of the leading strand is initiated by telomeric repeat binding factor 2 (TRF2, *Section 1.1.2.1*) to produce the corresponding G-tail (Arnoult and Karlseder 2015) (**Figure 1.2**).

As a result of incomplete lagging strand replication, telomeres shorten with each cell division in somatic cells that do not express telomerase. As such, telomeric DNA is lost following each cell division instead of genomic coding DNA. In human somatic cells, telomeres shorten by 50-100 bp per population doubling (PD) (Baird et al. 2003). Eventually, the accumulation of a few critically short telomeres will result in the initiation of the DNA damage response (DDR) via ataxia telangiectasia mutated (ATM) or ataxia telangiectasia and Rad3 related (ATR) signalling (Hemann et al. 2001). This triggers p53-driven replicative senescence, during which cell division ceases and cells may be arrested in G<sub>1</sub>/S-transition (Artandi and Attardi 2005). The number of times a normal somatic cell can undergo division before entering senescence is referred to as the Hayflick limit (Hayflick 1965).

Therefore, the replicative lifespan of the cell is strongly-influenced by telomere length. Telomere shortening is thought to act as a tumour suppressive mechanism as it protects against the propagation of cells which have acquired mutation and DNA damage over time (Shay and Wright 2000). However, telomere length-independent senescence also occurs as a result of telomere damage (Vitorelli and Passos 2017).



**Figure 1.2:** During DNA replication, the leading strand is replicated 5' to 3' by a DNA polymerase. The G-rich overhang is generated by resectioning. RNA primers (green) are used to generate Okazaki fragments in the 5' to 3' direction for the lagging strand. RNA primers are then degraded and replaced by DNA. However, Okazaki fragments cannot be generated at the position of the terminal primer. This creates a gap which eventually becomes the G-rich overhang.

### 1.1.2.1 Shelterin

Shelterin is fundamental for telomere function as without telomeric protection, aberrant genetic rearrangement would occur. Hundreds of shelterin complexes or telosomes are found at every telomere. In humans, the shelterin complex is composed of six different proteins:

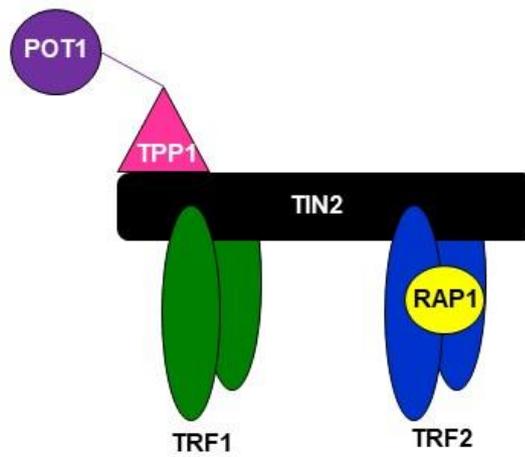
- Telomeric repeat binding factor 1 (TRF1)
- Telomeric repeat binding factor 2 (TRF2)
- TRF1- and TRF2-interacting nuclear factor 2 (TIN2)
- Protection of telomeres 1 (POT1)
- Repressor/activator protein 1 (RAP1)
- TINT1-PTOP-PIP1 (TPP1) (**Figure 1.3A**)

These telomere-associated proteins 'cap' telomeric DNA and result in the formation of a T-loop structure (Griffith et al. 1999) which protects the ends of chromosomes from recognition as sites of DNA damage (Liu et al. 2004) (**Figure 1.3B**). The G-tail folds back and inserts itself within the upstream double-stranded telomere sequence. This is mediated by TRF1 which is able to bend telomeric DNA at a 120° angle (Bianchi et al. 1997). This results in the displacement of a strand of the double-stranded telomere sequence, referred to as the D-loop. This D-loop structure is stabilised by POT1 (Lei et al. 2004; Loayza et al. 2004).

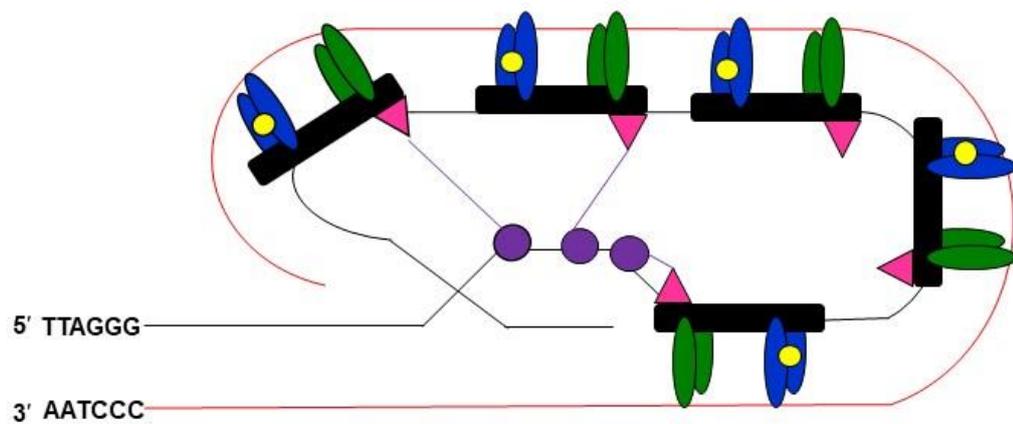
The components of the shelterin complex are therefore important for regulating telomere length by modulating the access of telomerase to telomeric DNA. In addition, the T-loop structure formed is essential for ensuring that telomeric ends are not identified by DNA damage apparatus as double-strand breaks (DSBs) (Doksani and de Lange 2014). As such, shelterin components are required to inhibit DNA damage response (DDR) signalling via the ATM or the ATR pathways. Activation of these signalling pathways triggers DNA repair via non-homologous end joining (NHEJ)

(Herbig et al. 2004) and homologous recombination (HR) (Thanasoula et al. 2012) at chromosome ends, respectively.

A



B



**Figure 1.3:** (A) The shelterin complex is composed of six different proteins, TRF1, TRF2, TIN2, TPP1, POT1 and RAP1. (B) The T-loop structure formed at telomeric DNA is stabilised by the shelterin complex.

### 1.1.3 Telomerase

Telomerase is a ribonucleoprotein that elongates telomeres by reverse transcription, using an internal RNA primer to add TTAGGG repeats to the 3' chromosomal terminus (Greider and Blackburn 1985). Telomerase is expressed in germline and stem cells, enabling these cells to maintain their telomere lengths. However, telomere expression in somatic cells is low or undetectable resulting in telomere shortening following cell division (Harley et al. 1990). Although telomerase is not expressed in somatic cells, telomerase reactivation/upregulation has been identified in ~90% of cancers resulting in cellular immortalisation (Kim et al. 1994).

The telomerase complex is comprised of a catalytic subunit, telomerase reverse transcriptase (TERT) and a telomerase RNA template component (TERC) subunit (Wenz et al. 2001). Telomerase is also associated with several proteins (dyskerin, NOP1, NHP2, TCAB1 and GAR1) that form a complex which confers stability and facilitates the accumulation of the TERC subunit in stem cells (Schmidt and Cech 2015).

TERT is made up of four conserved domains:

- Telomerase essential N-terminal (TEN) domain with affinity for telomeric ssDNA and TPP1
- Telomerase RNA binding (TRB) domain that interacts with TERC
- Reverse transcriptase (RT) domain responsible for catalytic activity
- C-terminal extension (CTE) domain

The TERC subunit contains an RNA templated sequence (3'-AAUCCC-5') which can be divided into alignment and sequencing sections. The alignment section binds to the telomeric repeats forming a DNA/RNA hybrid. The sequencing section is used as a template to reverse transcribe 5'-GGTTAG-3' repeats onto the 3'-end of the telomere which is catalysed by the RT domain. The RNA template then dissociates

from the telomere and translocates along the telomere to repeat the process (Blackburn and Collins 2011).

Whilst components of the shelterin complex modulate the ability of telomerase to access telomeric DNA, telomerase activity is regulated at multiple levels. For example, telomere elongation is only observed during S-phase of the cell cycle (Chen and Lingner 2013). Following S-phase, the DNA replication complex CST (CTC1, STN1 and TEN1) blocks access of telomerase to telomeric DNA (Chen and Lingner 2013).

#### **1.1.4 Alternative lengthening of telomeres (ALT)**

A telomerase-independent alternative telomere lengthening mechanism (ALT) has been identified in ~10-15% of cancers, generally of mesenchymal origin (Lafferty-Whyte et al. 2009), that maintain telomere length in the absence of telomerase (Bryan et al. 1997). These cancers are associated with a particularly poor prognosis (Dilley and Greenberg 2015). ALT-positive cells are associated with telomeric DNA that has separated from chromosomal ends (Cesare and Griffith 2004). This extrachromosomal telomeric DNA can be circular or linear. ALT is characterised by the presence of C-circles, which are comprised of partially single-stranded circles of telomeric DNA in which the C-rich strand is intact. Replication fork collapse induced by DSBs at telomeres that are not rescued result in the formation of these C-circles (Zhang et al. 2019).

It is suggested that ALT relies on HR and involves the extension of the 3' telomeric overhang which invades other telomeric DNA, which is used as a template for replication (Henson et al. 2002). The telomeric DNA used as a template may be:

- from a sister chromatid
- part of the same telomere (through telomere-loop formation)
- from the telomere at a different chromosome end

- from other extrachromosomal telomeric DNA

Whilst it has been demonstrated that telomeres within ALT cells are able to initiate a DDR, telomere fusions are repressed in these cells. It is therefore suggested that this intermediate telomere state (between uncapped and fully capped) is able to facilitate recombination-mediated telomere replication (Cesare and Reddel 2010) rather than NHEJ-mediated fusion.

### 1.1.5 Telomere-Position Effect (TPE)

The Telomere-Position Effect (TPE) has been identified in human cells and refers to the telomere length-dependent epigenetic silencing of telomere-proximal genes (Baur et al. 2001). It is hypothesised that increasing telomere lengths are associated with the 'spread' of heterochromatin towards the subtelomere. As heterochromatin is more compact, this results in the silencing of telomere-proximal genes. In contrast, shorter telomere lengths are associated with a more open chromatin structure that does not inhibit expression (Doheny et al. 2008). In addition, some research suggests that when telomere length decreases significantly, TPE is responsible for the transcription of the human telomerase gene, *hTERT*, located at the 5p subtelomere. This would involve TPE over long distances (TPE-OLD), a mechanism by which telomere length can modulate the expression of genes which are up to 10 megabases (Mb) away via telomere looping (Robin et al. 2014). It has been demonstrated that longer telomeres result in the co-localisation of the *hTERT* locus with the 5p subtelomeric region, possibly mediated by TRF2 and telomere repeat-containing RNA (TERRA), which represses *hTERT* expression. Telomere shortening caused the dissociation of the *hTERT* locus from the subtelomere, leading to a change in chromatin structure which was associated with increased *hTERT* expression and subsequent telomere elongation (Kim et al. 2016).

### **1.1.6 Telomere Repeat-containing RNA (TERRA)**

Subtelomeric DNA is transcribed into a long non-coding RNA, Telomere Repeat-containing RNA (TERRA) (Luke and Lingner 2009). It is understood that TERRA regulates telomerase activity by binding to the RNA template (*Section 1.1.3*) (Lalonde and Chartrand 2020). In addition, it has been shown that TERRA facilitates the displacement of Replication Protein A (RPA) and the subsequent binding of POT1 to ssDNA (Flynn et al. 2011). It is also suggested that TERRA participates in the formation of heterochromatin at telomeric DNA (Deng et al. 2009).

## **1.2 MEASURING TELOMERE LENGTH AND IDENTIFYING FUSION EVENTS**

As telomeres are repetitive sequences, accurately measuring telomere length has proved challenging. Techniques have been developed which differ in accuracy and specificity, with some methods measuring telomere lengths at single chromosome ends and others measuring the mean telomere length at all chromosome ends.

### **1.2.1 Terminal Restriction Fragment (TRF) Analysis**

TRF was the first method developed for measuring telomere length. Genomic DNA is digested using restriction enzymes that frequently cleave different sites throughout the genome whilst leaving subtelomeric and telomeric regions intact (Kimura et al. 2010). Telomeres are then resolved using agarose gel electrophoresis. Following Southern blotting, a telomere-specific labelled probe is used to hybridise to the telomere repeats. The labelled telomeric DNA produces a smear due to the difference in telomere length between various chromosome ends. Therefore, the average telomere length is estimated based on comparisons to standardised DNA ladders (Mender and Shay 2015).

There are numerous disadvantages associated with this technique. As this method measures telomere lengths at all chromosome ends, the estimated average length is not accurate enough to detect subtle changes in telomere lengths or changes at specific chromosome ends. In addition, large amounts of DNA are required (>1 µg) and the quantity of probe binding is proportional to telomere length, meaning shorter telomeres produce lower signal so are harder to detect.

### **1.2.2 Quantitative Polymerase Chain Reaction (Q-PCR)**

Q-PCR requires use of a PCR machine which is able to perform PCR whilst detecting fluorescence emitted by fluorophores. Fluorescent probes are added to the PCR

reaction so that upon DNA binding, a fluorescent signal is recorded. The greater the amount of DNA amplified, the greater the fluorescent signal intensity.

Primer pairs amplify the telomeric sequence by targeting C- and G-rich sequences but the primers have incorporated mismatches to prevent dimerization. The longer the telomere, the greater the number of binding sites for telomere-specific fluorescent probes and therefore a greater signal intensity is recorded. For quantification of telomere length, a single-copy gene is also amplified. The ratio of telomere-specific signal compared with single-copy gene signal is calculated (Cawthon 2002). As it was demonstrated that pipetting errors could result in significant errors, an adapted technique, monochrome multiplex quantitative PCR (MMqPCR) was developed which involved amplifying the telomeric and single-copy gene template within the same well (Cawthon 2009). A key advantage of this technique is that it is high-throughput and thus large cohorts can be analysed for epidemiological studies (Aviv et al. 2011). Although these methods require less DNA than TRF analysis, they exhibit a high measurement error that limits its use for clinical applications (Aviv et al. 2011).

### **1.2.3 Fluorescence *in Situ* Hybridisation**

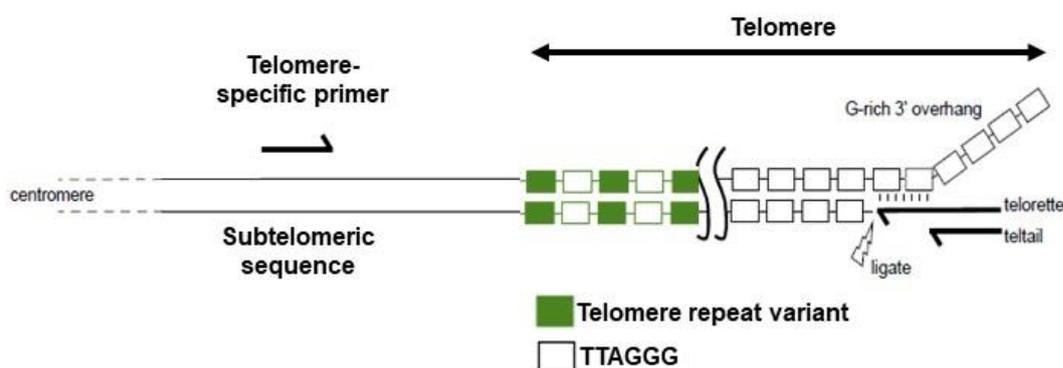
Quantitative fluorescence *in situ* hybridisation (Q-FISH) allows the visualisation of individual telomeres of metaphase chromosomes by using fluorescently-labelled probes. A fluorescence microscope is used to visualise chromosomes and the average telomere length is estimated by comparing the intensity of the signal with known standards (Lansdorp et al. 1996). Whilst this technique allows for the detection of short telomeres (200 bp) and telomere lengths at specific chromosome ends, mitotically active cells are required. Therefore, this technique cannot be used to measure telomere length for senescent or non-replicating cells.

Flow-FISH is the combination of Q-FISH and flow cytometry for the analysis of cells that are in suspension. The cells are treated with a fluorescently-labelled probe and then analysed using flow cytometry.

As the fluorescently-labelled probe used in all variations of FISH cannot bind to telomere repeats within the most proximal end of the telomere, the accuracy of this method for measuring short telomere lengths is reduced (Aubert et al. 2012). This technology exhibits a low measurement error and has thus been adopted for clinical diagnostic applications (Alder et al. 2018).

#### 1.2.4 Single Telomere Length Analysis (STELA)

STELA is a single-molecule long-range PCR-based assay which allows individual telomeres from specific chromosomal ends to be measured (Baird et al. 2003). The first step involves the annealing of the 'telorette' linker which has six bases of TTAGGG homology followed by a 20-nucleotide non-complementary tail to the G-rich 3' overhang of the telomere. The telorette linker is ligated to the 5' end of the complementary strand of the chromosome. This results in the end of the telomere being tagged with a non-complementary telorette tail. PCR is performed using a chromosome specific primer which binds to a unique sequence within the subtelomere alongside a 'teltail' primer that is identical to the 'telorette' tail (**Figure 1.4**).

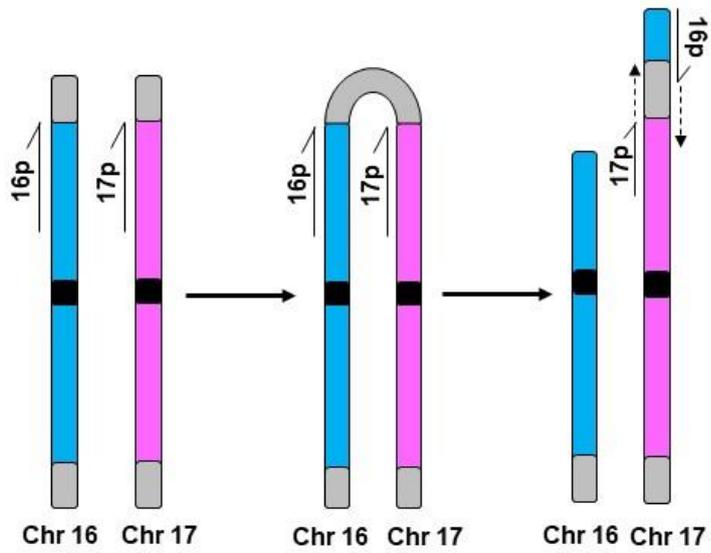


**Figure 1.4:** A telomere-specific primer is used to bind to a sequence within the subtelomere. A second telorette linker is used which binds at the end of the telomere via complementary base pairing. It is then ligated to the telomere. The teltail primer is used in the PCR to bind to the telorette tail. Figure adapted from Baird et al. 2003.

STELA can be used to amplify telomeres at 2p, 9p, 11q, 12q, 16q, 17p, 18q and XpYp chromosome ends (Britt-Compton and Baird 2006). This technique allows the detection of extremely short telomeres, although it is not as effective for measuring very long telomeres (>20 kb). The sensitivity of this technique ensures that a very accurate measurement of telomere length can be achieved (Baird et al. 2003). Additionally, STELA can be performed with small quantities (picograms) of high-quality genomic DNA. However, STELA can only be used to analyse intact telomeres at chromosomal ends for which a specific subtelomeric primer can be designed and it is labour intensive which limits its applications to large cohorts and its clinical utility.

### 1.2.5 Detecting telomere fusions

Telomere fusions in human cells can be detected using FISH in metaphase chromosomes (Ducray et al. 1999) or using a single-molecule telomere fusion PCR assay (Capper et al. 2007). The telomere fusion assay involves the use of single orientation chromosome specific primers in the subtelomeric region. Fusion events are amplified when fusion primers are in close proximity. PCR products are resolved using agarose gel electrophoresis and following Southern blotting, hybridisation with chromosome-specific probes reveals which telomeres are involved in the fusion event (**Figure 1.5**). A key advantage of fusion-PCR is that events can be characterised at the molecular level as fusion amplicons can be purified for sanger sequencing or specialised paired-end high-throughput sequencing. This allows high-throughput sequence analysis of telomere fusion events for corroboration of the events and the fusion junctions (Liddiard et al. 2016).



**Figure 1.5:** Cartoon representation of chromosome 16 and 17 with dysfunctional telomeres that undergo inter-chromosomal fusion and breakage. Dotted lines show how 16p and 17p fusion primers amplify the fusion event.

### 1.3 TELOMERE DYSFUNCTION AND CANCER

Telomeres are crucial for regulating the replicative lifespan of somatic cells and protecting the ends of chromosomes from DSB repair. Telomere function is therefore fundamental for maintaining genomic integrity (O'Sullivan and Karlseder 2010).

There are a number of ways in which telomeric DNA can become damaged. For example, oxidative damage and UV-induced damage can induce DSBs (Von Zglinicki 2002). Similarly, critically short telomeres which cannot bind shelterin components may become substrates for DNA repair as they are mistakenly recognised as DSBs (Doksani and de Lange 2014). In addition, disruption of shelterin components can trigger the DDR, irrespective of telomere length (van Steensel et al. 1998; Celli and De Lange 2005). Telomeric DNA can also become damaged following replicative stress (Sfeir et al. 2009) as stalled replication forks can collapse into DSBs (Muraki and Murnane 2017).

Broken DNA is able to fuse with other regions of damaged DNA in the genome, leading to rearrangement of the genome and chromosomal instability (Cannan and Pederson 2016). In human cells, there are mechanisms that are responsible for recognising and repairing DSBs including homologous recombination (HR) and non-homologous end-joining (NHEJ) (Her and Bunting 2018). However, these repair mechanisms do not distinguish between DSBs and the ends of chromosomes that are no longer protected by telomeres. Therefore, chromosomal end-end fusions can be observed at critically short, uncapped telomeres or when components of the shelterin complex are dysfunctional. As such, suppression of the DDR at chromosome ends is dependent on both telomere structure and the interaction of shelterin subunits with components of the ATM- and ATR-signalling DDR pathways (Sfeir et al. 2010). For example, TRF2 and POT1 modulate ATM and ATR-signalling pathways thereby repressing DNA repair mechanisms such as HR and NHEJ. Depletion of either TRF2 or POT1 results in telomere deprotection and telomere fusion events mediated by classical non-homologous end-joining (C-NHEJ)

irrespective of telomere length (Celli and De Lange 2005). Likewise, TRF2 suppresses ATM-signalling which would otherwise initiate C-NHEJ.

### **1.3.1 Telomere-driven crisis**

When telomeres become critically short in cells that lack fully functional cell-cycle checkpoints, cells may be able to continue dividing beyond the Hayflick limit which eliminates the protective capacity of the telomere (Wright and Shay 1992). Recognition of uncapped, critically short telomeres as DSBs triggers a DDR which results in the inappropriate repair of telomeric DNA due to the loss the shelterin components that normally suppress NHEJ (Counter et al. 1992). This is referred to as telomere-driven crisis (Capper et al. 2007). Aberrant repair of critically short telomeres during crisis can lead to the formation of unstable dicentric chromosomes that break during cell division. This can lead to breakage-fusion-bridge (BFB) cycles as these chromosomes may be pulled in opposite directions during anaphase which causes them to break. Subsequent repair therefore leads to additional cycles of breakage and fusions, ultimately resulting in chromosomal rearrangements (Murnane and Sabatier 2004).

Telomere fusions that occur during telomere-driven crisis therefore contribute to genome instability which can cause mitotic arrest and autophagic cell death (Nassour et al. 2019). However, these fusion events may precipitate recombinations that facilitate the upregulation of telomerase and the escape from crisis. It is thought that telomere fusions are highly deleterious (Rybanska-Spaeder et al. 2014) as they can impede normal chromosome segregation (Stewénius et al. 2007) and lead to genomic catastrophe (Umbreit et al. 2019). These fusions also result in aneuploidy which promotes oncogenesis by increasing genetic heterogeneity and promoting tumour evolution (Ben-David and Amon 2020).

Telomere fusions associated with telomere-driven crisis have been detected in a variety of human cancers including chronic lymphocyte leukaemia, breast, pancreatic

and colon cancers (Lin et al. 2010; Tanaka et al. 2012; Roger et al. 2013; Tanaka et al. 2014; Hata et al. 2018).

The upregulation of telomerase in several cancer types has been linked to changes associated with *hTERT* (Leao et al. 2018). These changes include point mutations in the *TERT* promoter (Lazzerini-Denchi and Sfeir 2016), gene amplification (Jones et al. 2014) or other genomic rearrangements that result in the juxtaposition of *TERT* close to active enhancers (Dwight et al. 2018).

### 1.3.2 DSB repair pathways

The DNA repair pathway employed at critically short telomeres mistakenly recognised as DSBs may be dictated by the cell cycle. It has been demonstrated that telomere fusion events during telomere-driven crisis can be facilitated by HR (Mao et al. 2016) as well as classical and alternative non-homologous end joining (C-NHEJ and A-NHEJ). It is thought that the relative activities of these NHEJ repair pathways influence the ability of cells to escape crisis (Jones et al. 2014).

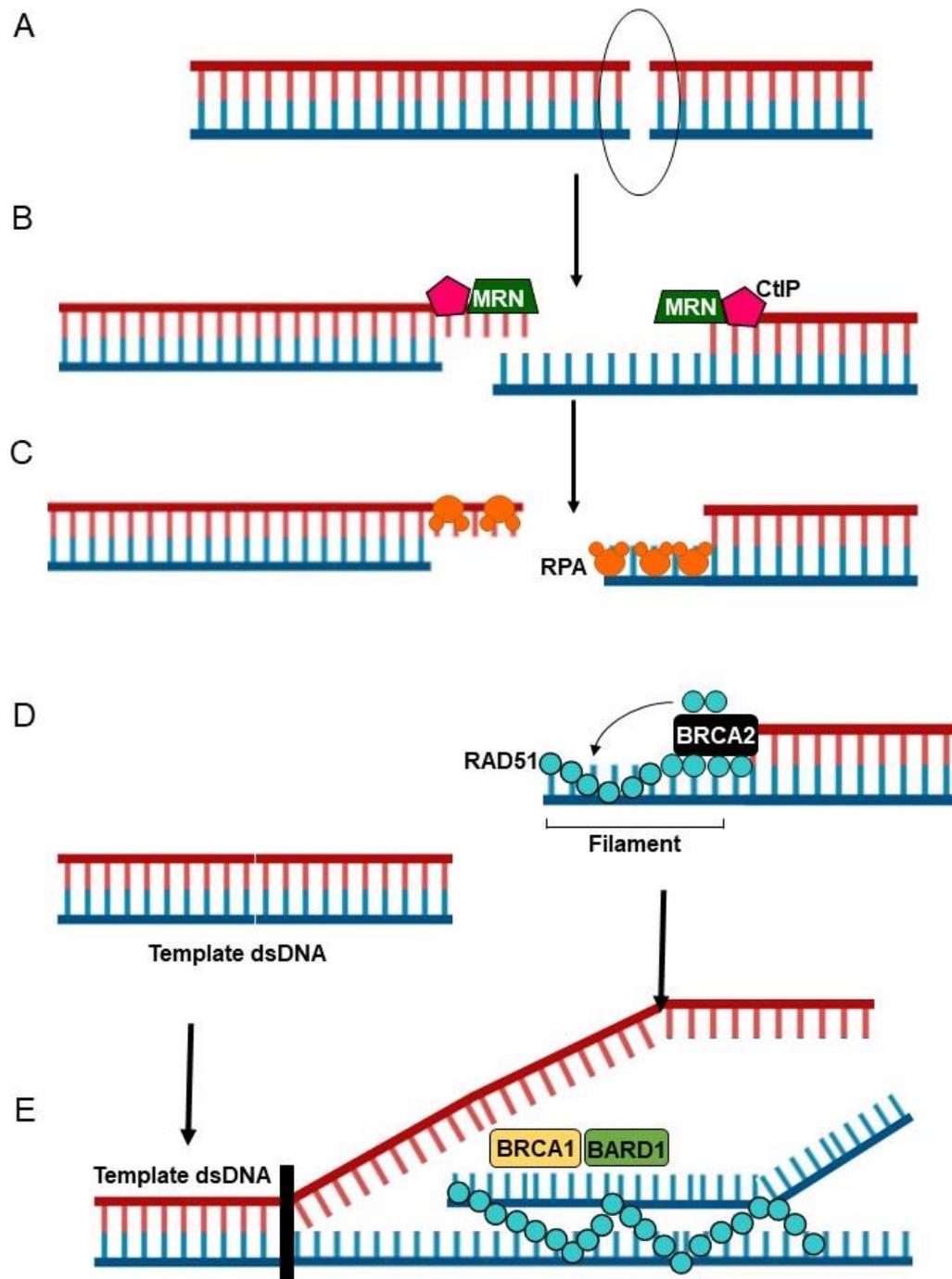
#### 1.3.2.1 Homologous Recombination (HR)

HR is an accurate form of repair that utilises sister chromatids as templates for repair. This mechanism is employed when DSBs are generated during S and G<sub>2</sub>-phases of the cell cycle when sister chromatids are available. Following a DSB, the MRE11-RAD50-NBS1 (MRN) complex and the CtBP-interacting protein (CtIP) localise to DSBs and resect broken ends in the 5' to 3' direction via nucleolytic degradation (Liu and Huang 2016). This reveals microhomology and results in the formation of 3' ssDNA which provides a platform for the recruitment of other proteins involved in repair. RPA binds to and stabilises the ssDNA, following which Breast Cancer type 2 susceptibility protein (BRCA2) facilitates the exchange of RPA for RAD51 on ssDNA (Jensen et al. 2010). This RAD51-ssDNA filament invades a homologous DNA

molecule, catalysed by the BRCA1-BARD1 complex. This homologous template is used to correctly repair the DSB and replace any missing sequence (Zhao et al. 2017) **(Figure 1.6)**.

### **1.3.2.2 HR at telomeres**

Previous studies have demonstrated an active role for HR at telomeric DNA. One study (Mao et al. 2016) involved following the fate of different human cell types in which subtelomeric and telomeric DSBs had been induced. Following the generation of DSBs within telomeric DNA, it was revealed that these telomeres formed physically proximate clusters which were associated with the preferential localisation of 53BP1 foci (53-binding protein 1), which is an important regulator of the cellular response to DSBs (Panier and Boulton 2013). In addition, the abrogation of RAD51 by exposure to B02, a specific inhibitor of HR, caused an inhibition of telomeric DSB repair (Mao et al. 2016). Altogether, these data suggest that HR is involved in the repair of telomeric DSBs. As HR is dependent on resection, it is likely that telomere fusions can only be mediated by HR during S/G<sub>2</sub>-phases of the cell cycle, as previously discussed.



**Figure 1.6:** (A) DSBs can be repaired by HR. (B) MRN-CtIP are recruited to the break and generate ssDNA via nucleolytic processing. (C) The ss-overhangs are coated and stabilised by RPA. (D) BRCA2 catalyses the replacement of RPA with RAD51. (E) The RAD51-ssDNA filament invades a homologous DNA molecule, catalysed by the BRCA1-BARD1 complex.

### **1.3.2.3 Break-Induced Replication (BIR)**

BIR is a form of HR that repairs DNA breaks that only have one end, so contributes to the repair of broken replication forks (Malkova and Ira 2013). BIR can also efficiently repair DSBs that result in only one end which displays homology to the genome, or when both ends display homology, but in different places (Lydeard et al. 2007). In addition, it has been suggested that nonallelic BIR may lead to translocations and BIR which is initiated at sites displaying microhomology can result in copy number variations (CNVs) and complex chromosomal rearrangements (Hastings et al. 2009). Furthermore, it has been demonstrated that DNA single-end substrates can form at chromosome ends in telomerase-deficient cells or upon the mis-segregation of fragmented chromosomes (Malkova and Ira 2013). Therefore, this repair pathway may be utilised during telomere-driven crisis as cells may be subject to increased replication stress and telomere fusion events. In addition, it is postulated that a RAD52-dependent and a RAD52-independent BIR pathway enables ALT (Zhang et al. 2019).

### **1.3.2.4 Single-strand Annealing (SSA)**

SSA also requires end resection that generates 3' ssDNA and reveals homologous sequences that flank the break site (Bhargava et al. 2016). RPA binds to ssDNA tails which initiates ATR-signalling and facilitates repair. The homologous sequences that flank the break sites are annealed by RAD52, forming a synapsed intermediate which is then subject to endonucleolytic cleavage of the 3' ssDNA tails by ERCC1 which forms a complex with XPF (Faridounnia et al. 2018). RAD52 enhances the nuclease activity of the ERCC1/XPF complex (Motycka et al. 2004). The requirement of these components therefore distinguishes this repair mechanism from both HR and A-NHEJ. The end resection required to reveal homologous sequences results in large deletions, therefore SSA results in relatively mutagenic repair (Bhargava et al. 2016). It has been demonstrated that SSA represents an essential pathway in ovarian

cancers which are associated with HR-dysfunction. As SSA is an error-prone repair mechanism, it is postulated that the increased use of this pathway in ovarian cancer results in genomic instabilities (Deniz et al. 2017).

### **1.3.2.5 Classical Non-Homologous End-Joining (C-NHEJ)**

C-NHEJ is active throughout the cell cycle and the predominant form of repair in non-dividing cells. This repair mechanism can result in minimal sequence alterations at repair junctions. C-NHEJ is initiated when the Ku70-Ku80 (Ku) heterodimer binds to the broken ends with high affinity. Ku is then able to recruit the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) which phosphorylates a number of other proteins involved in the repair including Artemis, a terminal end-processing enzyme that cleaves ssDNA overhangs (Ma et al. 2005). The scaffold protein, X-ray Repair Cross-Complementing protein 4 (XRCC4) associates with DNA ligase 4 (LIG4) which catalyses the ligation of the broken ends (Nick McElhinny et al. 2000).

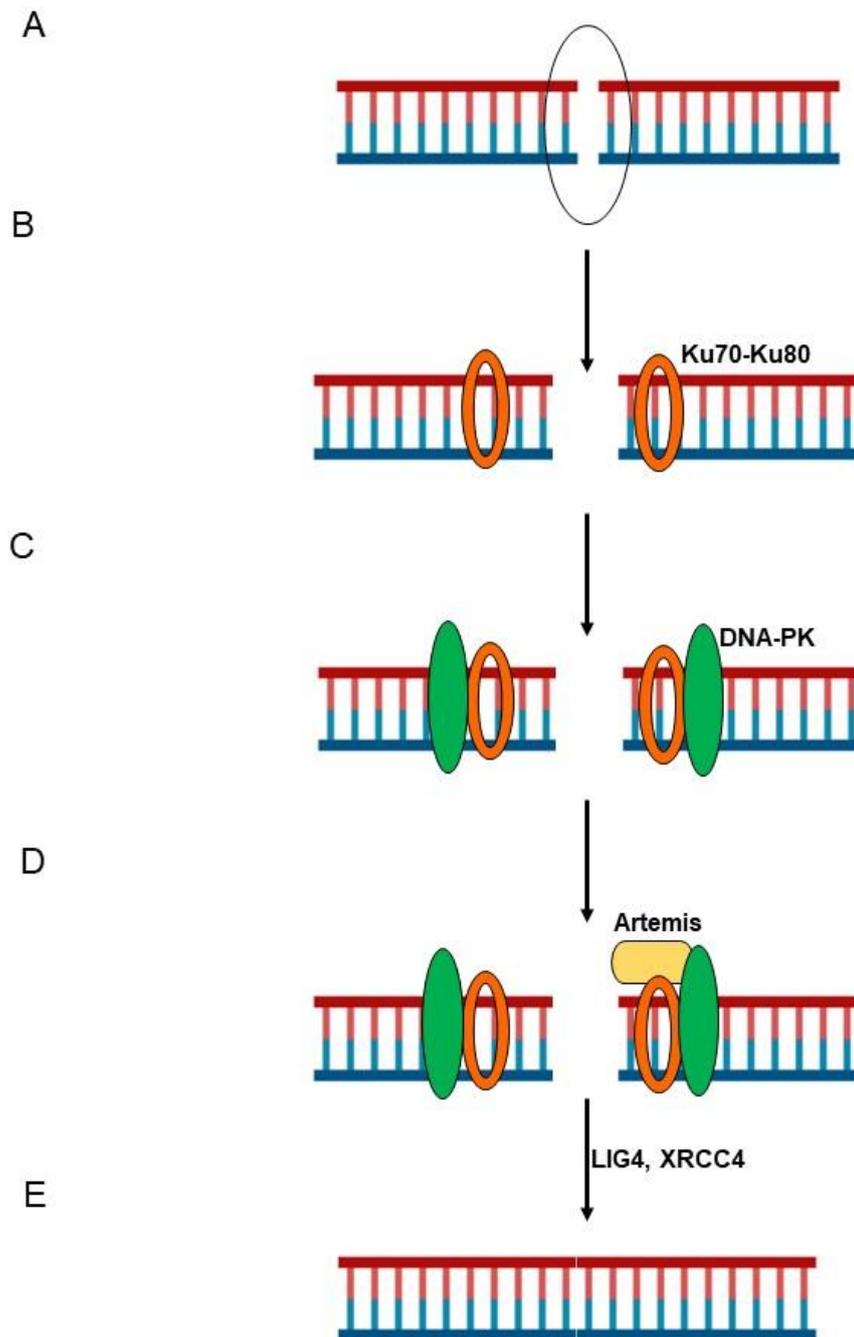
C-NHEJ is a flexible repair process that can employ different proteins depending on the complexity of the break. For example, if broken ends are incompatible, DNA polymerase mu and lambda (Pol  $\mu$  and Pol  $\lambda$ ) can perform template-dependent/independent DNA synthesis (Chang et al. 2016; Chang et al. 2017)

**(Figure 1.7)**

### **1.3.2.6 C-NHEJ at telomeres**

C-NHEJ is repressed at telomeres by TRF2, demonstrated by the formation of telomere fusions following disruption of TRF2. These telomere fusions had a mutational profile that was typical of C-NHEJ-mediated repair (Celli and De Lange 2005). It has also been demonstrated that inter-chromosomal telomere fusions mediated by LIG4-dependent C-NHEJ occur during telomere-driven crisis (Jones et al. 2014). In addition to chromosome end-end fusions, critically short telomeres are

able to fuse with non-telomeric loci, particularly within coding regions (Liddiard et al. 2016). These genomic telomere fusion events are mainly attributed to C-NHEJ and if telomerase is reactivated, the resulting gene disruption may cause detrimental phenotypic changes or facilitate malignant transformation.



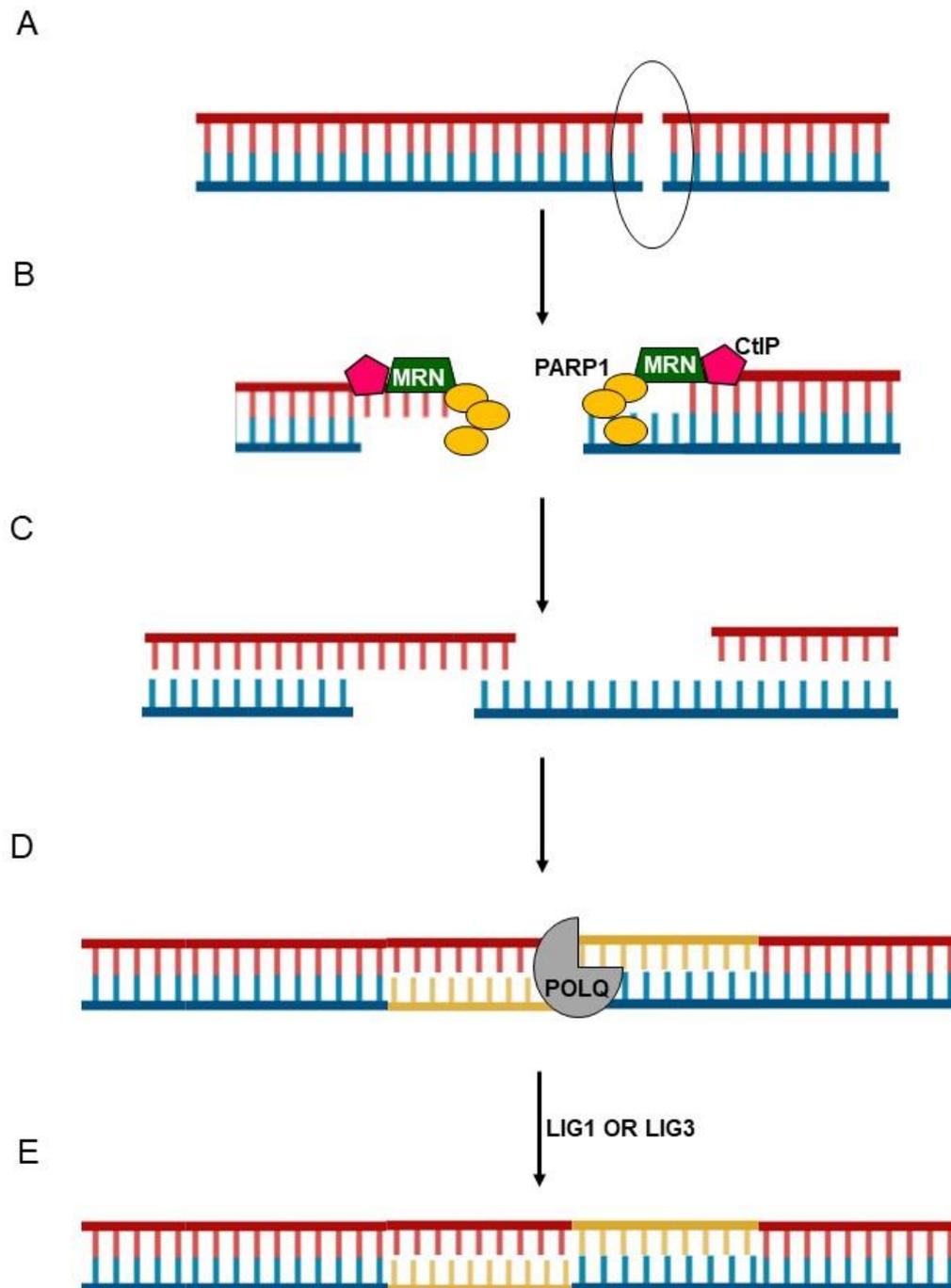
**Figure 1.7:** (A) DSBs can be repaired by C-NHEJ. (B) The Ku heterodimer is recruited to the break and binds with high affinity to DNA. (C) Ku recruits DNA-PK which phosphorylates downstream targets including (D) Artemis which cleaves ss-overhangs. (E) Xrcc4 associates with Lig4 which ligates broken ends.

### 1.3.2.7 Alternative Non-Homologous End-Joining (A-NHEJ)

A-NHEJ is often referred to as Microhomology-Mediated End-Joining (MMEJ) as this repair mechanism is characterised by microhomology (2-20 bp) as well as deletions and insertions which scars repair sites and occurs independently of the Ku heterodimer (Boulton and Jackson 1996). As with HR, A-NHEJ is most active during S and G<sub>2</sub>-phases of the cell cycle (Hustedt and Durocher 2016), and depends on resection by MRN and CtIP (Lamarche et al. 2010). However, A-NHEJ is mediated by Poly(ADP-ribose) polymerase 1 (PARP1) (Chaudhuri and Nussenzweig 2017). Resected ends undergo base pairing which initiates annealing and promotes synapsis of opposite ends of the DSB. DNA polymerase theta (Pol $\theta$ , POLQ) performs fill-in synthesis which stabilises the annealed intermediates and thus promotes end-joining via DNA ligase 3 (LIG3) or DNA ligase 1 (LIG1) (Liddiard et al. 2018). It is thought that the deletions are a result of the nucleolytic processing, whereas templated and non-templated insertions can be introduced by POLQ (**Figure 1.8**).

### 1.3.2.8 A-NHEJ at telomeres

It has been revealed that intra-chromosomal fusions that arise during telomere-driven crisis can be mediated by LIG3/LIG1-dependent A-NHEJ (Jones et al. 2014; Liddiard et al. 2018). These fusion events displayed increased microhomology and insertions at the fusion junctions compared with C-NHEJ-mediated inter-chromosomal fusions. Furthermore, it has been demonstrated that the escape from crisis is dependent on LIG3 (Jones et al. 2014). It is therefore postulated that C-NHEJ-mediated telomere fusions are not sufficient to mediate the escape from crisis, rather, A-NHEJ-mediated intra-chromosomal telomere fusion events are required to facilitate crisis escape.



**Figure 1.8:** (A) DSBs can be repaired by A-NHEJ which is facilitated by (B) PARP1 and requires resection by the MRN and CtIP. (C) Base pairing at resected ends drives annealing, promoting synapsis. (D) Annealed ends are stabilised by POLQ which performs fill-in synthesis and can introduce templated or non-templated insertions (yellow). (E) Ends are ligated by either LIG1 or LIG3.

### 1.3.3 DNA repair at telomeres

The DNA repair pathway employed to repair DSBs may be influenced by the type of break and the stage of the cell cycle during which it occurs. HR, A-NHEJ and SSA all require resection via the MRN complex which can only occur during S and G<sub>2</sub>-phases of the cell cycle. In contrast, C-NHEJ is active throughout the cell cycle, due to its capacity for resection-dependent/independent repair (Hustedt and Durocher 2016). It has been suggested that C-NHEJ is more prevalent than A-NHEJ as Ku binds with greater affinity to DNA than PARP1 (Wang et al. 2006). Furthermore, once the Ku heterodimer has bound to the DSB, resection is repressed, preventing the use of resection-dependent repair (Symington and Gautier 2011).

Dysfunctional/damaged telomeres can be repaired by both NHEJ pathways, resulting in the formation of inter- and intra-chromosomal telomere fusion events with varying proportions of microhomology, insertions and deletions (indel) at fusion junctions. As this indel mutational signature has also been detected at telomere fusion junctions in human cancers, the repair of critically short or damaged telomeres via NHEJ repair pathways may contribute to the chromosomal rearrangements observed in human malignancies (Hasty and Montagna 2014).

Failure to repair DSBs or aborted repair may result in the accumulation of DNA damage which triggers apoptosis (Hakem 2008).

## 1.4 DNA POLYMERASE THETA (POLQ)

DNA polymerases are essential for genome stability and maintenance due to their crucial roles in DNA replication and repair (Loeb and Monnat 2008). Therefore, the expression and activities of DNA polymerases are highly regulated. The human genome encodes at least 14 DNA polymerases which are involved in:

- Semi-conservative DNA replication (Pol  $\alpha$ , Pol  $\delta$ , Pol  $\epsilon$ )
- Mitochondrial DNA replication and repair (Pol  $\gamma$ )
- Base excision repair (BER) (Pol  $\beta$ , Pol  $\epsilon$ )
- NHEJ (Pol  $\lambda$ , Pol  $\mu$ , Pol  $\theta$ )
- Immunological diversity (Pol  $\lambda$ , Pol  $\mu$ )
- Translesion synthesis (Pol  $\zeta$ , Pol  $\theta$ )

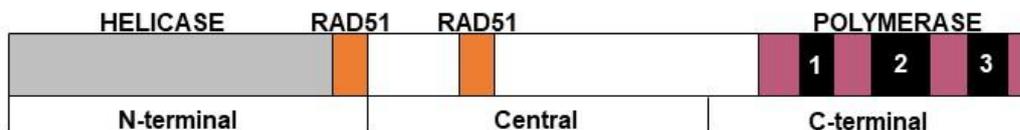
Some of these DNA polymerases perform more than one function and can be active in more than one DNA processing pathway (Garcia-Diaz and Bebenek 2007).

### 1.4.1 Discovery

Initial studies in *Drosophila melanogaster* carrying Mus308 (*POLQ* ortholog, DNA pol  $\theta$ ) mutations revealed that these mutants were hypersensitive to DNA damaging agents that cause DNA interstrand cross-links (ICL) such as hexamethylmelamine (Aguirrezabalaga et al. 1995). Further studies in mice showed that *POLQ* function was independent of ATM signalling (Shima et al. 2004) and research using mammalian cells indicated that *POLQ* expression confers resistance to DSBs induced by ionising radiation (Higgins et al. 2010). Additional studies in *Drosophila melanogaster* revealed that mus308 activity was independent of both C-NHEJ and HR and was associated with a distinct form of DNA repair which resulted in deletions, insertions and microhomology at repair junctions (Chan et al. 2010). Altogether, these data suggest that *POLQ* has an important role in DSB repair which is distinct from both C-NHEJ and HR.

### 1.4.2 POLQ structure

In 2003, human *POLQ* was purified and its gene product, POLQ, was characterised (Seki et al. 2003). It is now understood that human POLQ is a ~290 kDa (kilodaltons) protein that has a Hel308-type helicase domain in the N-terminal part of the protein, an unstructured central domain and an A-family polymerase domain at the C-terminus (**Figure 1.9**). Remarkably, this is currently the only known enzyme identified in higher eukaryotes that has both a helicase and polymerase domain. These helicase-polymerase proteins have been identified in bacteria, viruses and archaea and have roles in DNA repair and the initiation of DNA replication (Guilliam et al. 2015). Interestingly, there are RAD51 binding domains in the central and helicase domains. The domain arrangement of the POLQ protein is highly conserved in vertebrates (Yousefzadeh and Wood 2013).



**Figure 1.9:** Cartoon representation of human *POLQ*. The helicase domain is shown in grey, RAD51 binding domains in the helicase and central domains are shown in orange. The polymerase domain is shown in purple, with the three insertion loops highlighted in black.

#### 1.4.2.1 C-terminal polymerase domain

Altogether, the A family polymerase encoded by the C-terminal domain has three conserved insertion loops. Whilst this polymerase has similarities with Pol I type enzymes, this polymerase is unique due to insertion loop 2 which is essential for terminal transferase and translesion synthesis activity (Zahn et al. 2011). Currently, no other A family polymerases are known to be able to perform both terminal transferase and translesion synthesis activities.

### 1.4.2.2 N-terminal helicase domain

The superfamily 2 (SF2) helicase encoded by the N-terminal domain shares sequence homology with the mammalian helicase, Hel308 or HELQ, which functions in DNA replication repair (Richards et al. 2008). Biochemical studies have demonstrated that the helicase domain is able to efficiently unwind DNA with 3'-5' polarity (Ozdemir et al. 2018). Moreover, this study found that the POLQ helicase was able to unwind blunt ended DNA, and DNA with both 3' and 5' overhangs. In addition, it was shown that the POLQ helicase was able to unwind DNA-RNA hybrids as well as replication forks, exhibiting a preference for unwinding the lagging strand.

### 1.4.3 POLQ function

#### 1.4.3.1 Terminal transferase activity

It has been shown that POLQ facilitates the stabilisation of A-NHEJ intermediates by promoting DNA synapsis of 3' ssDNA overhangs. The opposing ss-overhang is used as a template to extend DNA which stabilises the A-NHEJ intermediate before the second overhang is extended. Whilst POLQ has previously been referred to as a 'promiscuous' (Hogg et al. 2012) enzyme, Kent T et al (2016) describe POLQ as 'among the most proficient terminal transferases known' (Kent et al. 2016). This cell-free study revealed that POLQ is able to oscillate between three different modes of terminal transferase activity without dissociating from the ssDNA:

1. Template-independent
2. Transient template-dependent extension in *cis* (snap-back replication)
3. Transient template-dependent extension in *trans*

It was suggested that template-independent terminal transferase activity is the predominant form of extension, but extension products could be generated by more than one mode of terminal transferase activity. This ability to alternate between

different types of terminal transferase may explain the combinations of templated and non-templated insertions observed at repair junctions following A-NHEJ.

In addition, Manganese ( $Mn^{2+}$ ) was identified as a co-factor which binds within the POLQ active site and promotes template-independent terminal transferase activity, resulting in the formation of longer extension products.  $Mn^{2+}$  was shown to promote this activity at low concentrations (0.2 mM) which are found in cells. Other studies have demonstrated that  $Mn^{2+}$  is a necessary co-factor for MRN, which is required for generating 3' overhangs during A-NHEJ (Truong et al. 2013).

Further studies have revealed that following induced chromosomal translocations in murine cells, mutant cells lacking a functioning POLQ helicase domain displayed a significant reduction in the frequency of chromosomal translocations (Mateos-Gomez et al. 2017). This was also observed for cells lacking a functioning POLQ polymerase domain. These results therefore suggest that the POLQ helicase domain is also crucial for A-NHEJ repair in mammalian cells. The presence of insertions and microhomologies at DNA repair sites typifies A-NHEJ. Therefore, fusion break points in these mouse cells were also analysed and revealed a reduction in microhomology and insertion usage for both mutant cell lines, suggesting that the reduction in the frequency of chromosomal translocations was due to a deficiency in A-NHEJ (Mateos-Gomez et al. 2017).

#### **1.4.3.2 Regulation of HR**

In addition to its essential role in A-NHEJ to facilitate DSB repair, a role for POLQ in negatively regulating HR was also reported. During HR, RPA binds to ssDNA to stabilise the structure and prevents annealing of homologous sequences which have been exposed following resection, thereby preventing A-NHEJ and ensuring a faithful repair mechanism is employed. RPA also stimulates additional resection by BLM-EXO1 which further stimulates repair via HR (Audebert et al. 2004).

However, studies in mouse cells have demonstrated that cells lacking a functioning POLQ helicase domain and cells lacking a functioning polymerase domain displayed significant increases in HR compared with cells with fully functional POLQ (Mateos-Gomez et al. 2017). This suggests that both polymerase and helicase activities are required for the repression of HR. As it had previously been suggested that sites of DNA damage that harbour ss-overhangs are preferentially repaired by either HR or A-NHEJ over C-NHEJ (Bothmer et al. 2017), this study also investigated the effect of POLQ helicase or polymerase depletion on HR-mediated repair of breaks with ss-overhangs. The significant increase in HR-directed repair observed for both POLQ mutants compared with controls conforms with previous data which suggests lesions containing ss-overhangs are preferentially repaired by HR or A-NHEJ. These data also further indicate that both the POLQ helicase and polymerase activities are required for the suppression of HR.

This study investigated whether the POLQ helicase promotes annealing of complementary ssDNA. The POLQ helicase stimulated annealing of complementary ssDNA and synapsis of ss-overhangs containing microhomologies in an ATP-independent manner in a cell-free system. Interestingly, in the presence of RPA, the POLQ helicase was also able to stimulate annealing in an ATP-dependent manner. It has been demonstrated that the POLQ helicase is able to form stable 400 kDa tetramers in solution (Newman et al. 2015). It is therefore suggested that this configuration enables the POLQ helicase to reduce the mobility of ssDNA compared to the RPA-ssDNA complex.

The ability of the POLQ helicase to actively displace RPA from ssDNA allows annealing and subsequent synapsis of the ss-overhangs, thereby ensuring the annealed intermediate is now a substrate for A-NHEJ.

In addition to facilitating the removal of RPA from ssDNA, it has been demonstrated that POLQ is able to interact directly with RAD51 via its intrinsic RAD51 binding

domains. One study revealed that depletion of POLQ caused a significant increase in basal and radiation (IR)-induced RAD51 foci (Ceccaldi et al. 2015). Furthermore, IR-induced RAD51 foci was fully reduced by full-length WT POLQ but not POLQ which had mutations in the ATPase catalytic residues or POLQ mutants lacking the RAD51 binding domains. However, POLQ mutants that lacked the polymerase domain caused a decrease in IR-induced RAD51 foci, suggesting that the disruption of RAD51 requires the N-terminal domain of POLQ. This study went on to examine POLQ function under replicative stress and showed that UV exposure caused an increase in RAD51 binding by POLQ, suggesting that POLQ regulates HR in cells under replicative stress.

Altogether, POLQ is able to actively displace RPA from ssDNA via its helicase domain and can interact directly with RAD51. Both interactions of POLQ with these components result in the disruption of HR and the upregulation of A-NHEJ.

#### **1.4.3.3 POLQ-mediated A-NHEJ at dysfunctional telomeres**

Studies in mouse cells have directly linked POLQ with A-NHEJ at deprotected telomeres (Mateos-Gomez et al. 2017). In this study, depletion of the shelterin complex in mouse embryonic fibroblasts (MEFs) was performed to initiate telomere fusions and telomere sister chromatid exchange (T-SCE) events. Depletion of RPA in these cells caused an increase in A-NHEJ and a significant reduction in the frequency of T-SCEs. This study further demonstrated that RPA expression in these cells prevented the increase in A-NHEJ.

In addition, human cells with a defective helicase domain were unable to restrict RPA accumulation at DSBs induced in telomere repeats (Mateos-Gomez et al. 2017). Taken together, these data suggest that both POLQ polymerase and helicase activities are essential for A-NHEJ and the suppression of HR. However, further

investigation of the role of POLQ in A-NHEJ at telomeric loci in human cells is required.

#### 1.4.3.4 Replication timing and cell cycle progression

It was hypothesised that POLQ might function during genomic replication due to its ability to replicate undamaged DNA - although it does so with low fidelity compared with other A family polymerases (Arana et al. 2008).

In one study, three clonal human cell lines were generated that stably expressed increasing amounts of POLQ compared with endogenous POLQ in controls. These cell lines were used to evaluate the impact of POLQ overexpression on cell proliferation. This study found that POLQ overexpression caused cells to accumulate in S-phase compared with controls. The cell line with the highest POLQ expression level showed accumulation of cells in G<sub>2</sub>/M-phase also, indicating that in the absence of external stress, POLQ expression affects cell cycle progression. In addition, it was also found that POLQ overexpression was associated with reduced DNA replication fork speed (Lemée et al. 2010).

Additional studies in human cell lines showed that under normal conditions, POLQ was recruited to chromatin with MCM (minichromosome maintenance) in early G<sub>1</sub>-phase of the cell cycle. Furthermore, it was found that POLQ was able to interact with ORC2 and ORC4 (Origin Recognition Complex), suggesting that POLQ may be able to bind to replication origins. These data suggest that POLQ may contribute to the formation of pre-replication complexes (pre-RC) at replication origins. The activation of replication origins is controlled in G<sub>1</sub>-phase by a specific timing programme referred to as 'timing decision point' (TDP) (Ryba et al. 2010). This study further revealed that POLQ plays a role in establishing this TDP. Additionally, human cells were generated which overexpress POLQ and were associated with delayed replication timing.

Taken together, these data suggest that POLQ regulates replication timing in particular regions of the genome and has a role during the cell cycle in unstressed cells (Fernandez-Vidal et al. 2014). This POLQ function is important as POLQ overexpression has been observed in colon, breast and lung cancers and is associated with poor clinical outcome (Kawamura et al. 2004; Pillaire et al. 2009;

Lemée et al. 2010). Furthermore, it has been postulated that genomic instability identified in haematologic malignancies may be as a consequence of abnormal temporal control of replication (Ryba et al. 2012). As such, high levels of POLQ in cancer cells could be involved in defining the modified temporal replication programme observed that might drive chromosomal instability (Bignell et al. 2010). Slowed replication allows for tolerance of DNA damage and suppresses genomic instability. Therefore, POLQ may have an important role in telomere replication and during telomere-driven crisis as cell proliferation is slowed. However, it is not currently understood how POLQ expression is regulated *in vivo* and the function of POLQ in non-replicating cells needs to be explored.

#### 1.4.4 POLQ and cancer

As POLQ has been implicated in chromosomal instability in *Drosophila melanogaster*, POLQ expression in human cancers was examined. 63% of non-small cell lung cancers (NSCLC) that were analysed (n=27) were shown to have elevated POLQ expression compared with normal tissue. In addition, increased POLQ expression was also observed for 39% of stomach (n=28) and 77% of colon carcinomas (n=26). POLQ was undetectable in most of the non-tumorous tissue from these samples indicating that POLQ is significantly and specifically upregulated in these tumours. Moreover, patients with high levels of POLQ expression had a poorer survival outcome. Therefore, it was hypothesised that POLQ overexpression may enable the survival or growth of cancer cells (Kawamura et al. 2004).

The survival of these POLQ-overexpressing cancer cells may be linked to the role of POLQ in modulating replication timing to accommodate DNA damage. POLQ may also mediate cell survival by promoting aberrant A-NHEJ-mediated DNA repair that gives rise to genomic instability which facilitates tumourigenesis. Alternatively, POLQ may enable cancer cell survival by providing resistance to cancer therapies.

##### 1.4.4.1 Breast cancer

As the deregulation of specialised DNA polymerases has been observed in neoplasia, Lemée et al (2010) investigated whether mis-regulation of DNA polymerases could be associated with the progression of breast cancer.

In this study, the expression of 13 DNA polymerases in 206 breast carcinomas were measured from two cohorts of patients. The first cohort revealed that POLQ was 3- to 26-fold higher in most of the tumours analysed and in 69.3% of the tumours analysed,  $\geq 5$ -fold increase in POLQ expression was observed. Furthermore, this upregulation of POLQ was significantly associated with adverse clinical outcome for the first cohort of patients. In addition, these observations were also supported by data obtained for a second, independent cohort of patients.

To determine whether POLQ actively contributes to tumour progression, POLQ-overexpressing human cells were generated to investigate the impact of POLQ expression on genomic stability. Cell lines overexpressing POLQ showed a significant increase in  $\gamma$ H2AX foci formation (which accumulates in cells with DNA damage) compared with controls. Further analysis revealed that the overexpression of POLQ results in the activation of the  $\gamma$ H2AX-ATM-CHK2 DNA damage checkpoint. In addition, a significant increase in the frequency of chromosomal abnormalities, mainly end-to-end fusions were detected for the POLQ-overexpressing cell lines compared with controls. Another independent study also observed POLQ overexpression in breast cancers and showed that POLQ upregulation was associated with increased rates of disease relapse (Higgins et al. 2010). This could be due to increased survival of cancer cells, facilitated by POLQ overexpression.

The observation that POLQ expression is upregulated in a variety of breast cancer tumours and is associated with poor clinical outcome suggests that this upregulation may not be related to particular carcinogens but instead may facilitate the survival of cancer cells. These data therefore indicate that POLQ could be a useful predictive marker for breast cancer patients thereby allowing identification of patients which require more aggressive therapies.

#### **1.4.4.2 HR-deficient tumours**

As with many other human cancers, POLQ overexpression has also been observed in epithelial ovarian cancers (EOCs) and is associated with adverse clinical outcome. As half of EOCs are HR-deficient due to mutations in HR-regulating genes such as BRCA1/BRCA2, these cancers are hyper-dependent on A-NHEJ (Ceccaldi et al. 2015). Previous studies have found that HR-deficient tumours are hypersensitive to POLQ, suggesting that A-NHEJ is essential for the survival of HR-deficient tumours. To investigate whether there was a synthetic lethality between HR genes and POLQ,

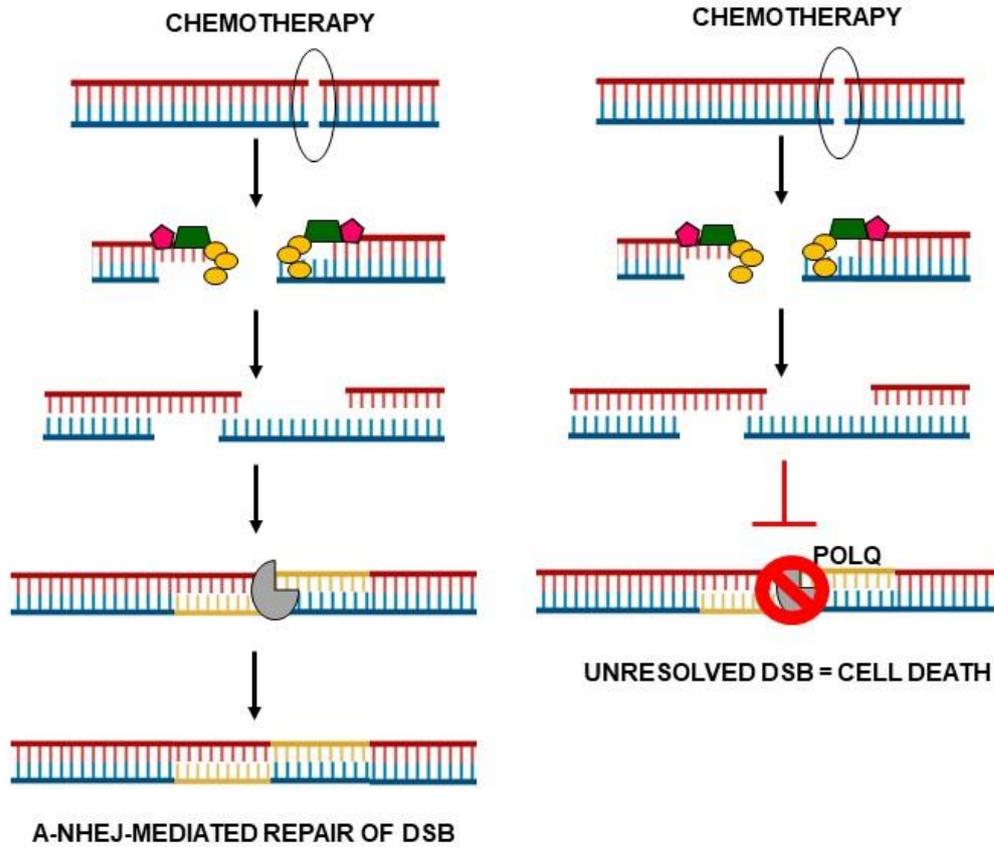
a HR-deficient ovarian cancer cell line was generated. This cell line was then subject to POLQ depletion and exposure to cytotoxic drugs to measure cell survival. POLQ depletion caused a reduction in the survival of these HR-deficient cells following exposure to a range of inhibitors providing additional evidence that HR-deficient cells are dependent on POLQ for survival. In addition, the loss of HR and POLQ in mice resulted in embryonic lethality (Ceccaldi et al. 2015).

Breast and ovarian cancers which are BRCA1/2-deficient are characterised by HR-deficiency and high levels of replication stress which leads to genomic instability (Pathania et al. 2014). In healthy cells, BRCA1/2 cooperate with Fanconi Anemia (FA) proteins to stabilise replication forks and maintain genomic integrity (Schlacher et al. 2011). It has previously been shown that BRCA1/2-deficient tumours upregulate FANCD2 which is required for replication fork protection where it protects DNA strands from excessive nucleolytic degradation (Schlacher et al. 2011; Schlacher et al. 2012). FANCD2 is then able to facilitate replication restart once DNA repair is complete. One study has demonstrated that FANCD2 promotes POLQ recruitment and A-NHEJ at sites of DNA damage (Kais et al. 2016). Furthermore, this study revealed that FANCD2 expression was highly correlated with POLQ expression and depletion of FANCD2 caused a reduction in the efficiency of A-NHEJ. Similarly, FANCD2 inhibition resulted in a decrease in the accumulation of POLQ at DNA break sites. As knockdown of POLQ did not affect FANCD2 foci formation, it was suggested that FANCD2 functions upstream of POLQ and facilitates the recruitment of POLQ to DSBs. Furthermore, FANCD2 overexpression in BRCA1/2-deficient cells confers resistance to PARP inhibitors through stabilization of replication forks.

#### **1.4.4.3 POLQ inhibitors**

It has been shown that POLQ is an essential component of the A-NHEJ pathway that facilitates inaccurate repair of DSBs. However, DSB-forming agents such as

chemotherapy are used for the treatment of tumours by inducing DSBs to initiate cell death. POLQ overexpression has been observed in EOCs which are deficient in HR and is associated with poor prognosis. Therefore, A-NHEJ is able to resolve DSBs induced by chemotherapy in these tumours. As a synthetic lethal relationship has been established between HR genes and POLQ, POLQ now represents a therapeutic target for the treatment of HR-deficient tumours. It has been suggested that the inhibition of POLQ for tumours carrying mutations in HR genes will result in more effective chemotherapy (**Figure 1.10**). However, more needs to be understood about the role of POLQ in healthy and cancerous human cells to determine whether POLQ inhibitors represent valid therapeutic treatment options.



**Figure 1.10:** Chemotherapy induces DSBs in HR-deficient tumours that are repaired via A-NHEJ resulting in cell survival (left). Targeted POLQ inhibition in HR-deficient tumours inhibits A-NHEJ, allowing the accumulation of unresolved DSBs leading to cell death (right).

## 1.5 AIMS OF THE PROJECT

Short dysfunctional telomeres and telomeres that are damaged as a result of replicative stress are able to undergo chromosomal fusion and rearrangement facilitated by end-joining repair pathways. Gross chromosomal rearrangements have been observed in cancer genomes and are associated with malignant progression (Lin et al. 2010). POLQ is a crucial component of the A-NHEJ pathway that mediates error-prone DSB repair. Furthermore, it has been shown that POLQ is overexpressed in many human cancers and this overexpression is associated with adverse clinical outcome (Lemée et al. 2010). As such, it is proposed that POLQ is a suitable target for therapeutic intervention for HR-deficient tumours. However, much more needs to be understood about the role of this enzyme in both healthy and cancerous, proliferating, and post-mitotic human cells.

This project had three broad aims which are dealt with in three results chapters:

1. **To determine the impact of POLQ expression on the cellular capacity to escape a telomere-driven crisis (Chapter 3).** To investigate this aim, telomere-driven crisis was induced in two independent human cell lines sufficient or deficient for POLQ.
2. **To investigate the role of POLQ in the NHEJ pathways utilised for telomere fusion (Chapter 4).** To address this question, telomere fusion events arising during telomere-driven crisis in the presence or absence of POLQ were analysed. In addition, DSBs were experimentally-induced in subtelomeric DNA to explore the role of POLQ in DNA damage responses.
3. **To assess how the loss of POLQ impacts the evolving cancer genome (Chapter 5).** To examine the genome-wide impact of POLQ-mediated repair, telomere fusion amplicons from WT and POLQ-deficient clonal populations progressing through telomere-driven crisis were

sequenced using paired-end Illumina sequencing. Moreover, whole genome sequencing was performed on these clones before and after telomere-driven crisis.

## Chapter 2 Materials and Methods

### **2.1 MATERIALS**

#### **2.1.1 Reagents**

Chemicals and reagents used were obtained from numerous sources including: Fisher, Invitrogen, Applied Biosystems, Thermo Scientific, New England Biolabs, Amersham biosciences/GE healthcare, Roche, Sigma-Aldrich, PerkinElmer, Bio-Rad, Abcam and Millipore.

#### **2.1.2 Plastic lab equipment**

Plastic and glass lab equipment used for experiments was obtained from Gilson, Becton Dickinson labware, Eppendorf, Thermo Scientific and SARSTEDT.

#### **2.1.3 Equipment/machinery**

Equipment used during this study was obtained from Bio-Rad, Applied biosystems, Thermo Scientific, Promega, Hybaid, Amersham biosciences and Qiagen.

## 2.2 METHODS

The Cardiff University Radiological Protection Course was completed and the use of P33 and P32 radioisotope was documented using IsoStock® software. Health and safety training, including COSHH assessments, were completed at University Hospital Wales, Cardiff.

### 2.2.1 Tissue Culture

#### 2.2.1.1 Cell lines and growth conditions

Three distinct wildtype (WT) human cell lines and cell lines in which *POLQ* had been disrupted via CRISPR/Cas9 editing were used for this study. A single clone of each of the cell lines was received.

HAP1 human near-haploid cells and HAP1 *POLQ* cells were received from Professor Geoffrey Higgins (Oxford University). This Human HAP1 parental control cell line was wild-type male (XY) HAP1 cells which were not isolated from one clone (pooled clones). As such they could be a mixture of haploid and diploid HAP1 cells. This WT cell line was isolated from KBM-7 leukemic cells which were then subject to serial subcloning. This cell line had a haploid karyotype except for a disomy of chromosome 8 (Kotecki et al. 1999). This cell line also harbours a translocation between chromosomes 9 and 22 known as the Philadelphia chromosome. In addition, a small portion of chromosome 15 is found in the middle of chromosome 19. The Philadelphia chromosome does not cause any duplication of genes, so this cell line is still considered haploid. However, the chromosome 15 fragment is a duplication so there are approximately 300 genes from chromosome 15 that are diploid (Nowell and Hungerford 1960; Kotecki et al. 1999). Therefore, this cell line is referred to as 'near-haploid'. The HAP1 *POLQ* cells were not derived from the same population as the HAP1 WT cell line. Any differences in how long these cell lines were cultured before receiving them is unknown.

Human primary retinal epithelial cells immortalised via hTERT expression, RPE1<sup>hTERT</sup> and RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup>, were obtained from Professor Agnel Sfeir (Skirball Institute / NYU). These female (XY) adherent cell lines are near diploid with a modal chromosome number of 46. These cell lines were derived by transfecting the RPE-340 cell line with the pGRN145 hTERT-expressing plasmid (ATCC® CRL-4000™). The RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup> cells were not derived from the same population as the RPE1<sup>hTERT</sup> WT cell line. Any differences in culturing before these clones were received is unknown.

Human Colon Carcinoma (HCT116) cells, HCT116 *POLQ*<sup>-/-</sup> and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cells were gifted from Professor Eric Hendrickson, (University of Minnesota). The stem line chromosome number was near diploid with the modal number at 45 (62%) and polyploids occurring at 6.8%. Q-band observations revealed the presence of the Y chromosome in 50% of cells (ATCC® CCL-247™). The HCT116 *POLQ*<sup>-/-</sup> cells were not derived from the same population as the HCT116 WT cell line. The duration for which these cell lines were cultured before receiving them is unknown.

A 2 bp deletion at the CRISPR/Cas9-targeting sequence in exon 1 of *POLQ* was induced to generate HAP1 *POLQ*<sup>-</sup> cells. The same 1 bp deletion in exon 2 at the CRISPR/Cas9-targeting sequence was observed in both HCT116 and RPE1<sup>hTERT</sup> cells (*Chapter 3*).

HAP1 cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM, Life Technologies) in 25 or 75 cm<sup>2</sup> flasks. Cells were incubated in a 37°C, 5% CO<sub>2</sub> incubator. IMDM was supplemented with 10% v/v Foetal Calf Serum (FCS, Thermo Fisher Scientific), 100 U/mL Penicillin (Sigma), 0.1 mg/mL Streptomycin (Sigma) and 2 mM L-glutamine (Sigma).

RPE1<sup>hTERT</sup> cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies) in 25 or 75cm<sup>2</sup> flasks. Cells were incubated in a 37°C, 5% CO<sub>2</sub>

incubator. DMEM was supplemented with 10% v/v Foetal Calf Serum (FCS, Thermo Fisher Scientific), 100 U/mL Penicillin (Sigma), 0.1 mg/mL Streptomycin (Sigma) and 2 mM L-glutamine (Sigma).

HCT116 cells were cultured in McCoy's 5A (Modified) Medium (Life Technologies) in 25 or 75cm<sup>2</sup> flasks. Cells were incubated in a 37°C, 5% CO<sub>2</sub> incubator. McCoy's was supplemented with 10% v/v Foetal Calf Serum (FCS, Thermo Fisher Scientific), 100 U/mL Penicillin (Sigma), 0.1 mg/mL Streptomycin (Sigma) and 2 mM L-glutamine (Sigma).

A Primovert microscope and Motic 5.0 megapixel (MP) camera (Zeiss) was used to examine and photograph live cells. Motic Live Imaging software was used for image capture.

### **2.2.1.2 Trypsinisation and Cell Counting**

For the HAP1, HCT116 and RPE1<sup>hTERT</sup> adherent cell lines, cells were detached from the culture flasks using trypsin. Media was removed from the flasks, for a 75 cm<sup>2</sup> flask, 2 ml of trypsin (1x, Thermo Fisher Scientific), a proteolytic enzyme, was washed over the cells and the flask incubated at 37°C for 5 min. Trypsin was inactivated by the addition of 8 mL of fresh medium. Cells were then harvested, and a sample taken for cell counting before centrifugation (175 x g for 5 min) prior to resuspension in fresh culture media.

To count and assess cell viability, 1.3 µL of solution 13 (Acridine Orange 80 µg/mL and DAPI 40 µg/mL, Chemometec) was added to a 25 µL suspension of cells. The NucleoCounter NC-3000 (Chemometec) was used to count cells as per the manufacturer's guidelines.

### 2.2.1.3 Cryopreservation and thawing of cells

Approximately  $1 \times 10^5$ – $1 \times 10^7$  cells were centrifuged for 5 min to form cell pellets before being resuspended in 1 mL of cryopreservation media (50% appropriate growth media, 25%, 15% FCS and 10% DMSO). Cells were stored in freezing ampoules (Thermo Scientific) which were then placed into a cryofreezing container (Mr. Frosty) and stored at  $-80^\circ\text{C}$  for 24 h before being stored long-term in liquid nitrogen. To ensure sample freezing at a controlled rate of  $1^\circ\text{C}/\text{min}$ , Nalgene Mr. Frosty freezing containers were filled with isopropyl alcohol.

Cells were removed from liquid nitrogen and rapidly thawed in a water bath at  $37^\circ\text{C}$ . Freezing media was removed by adding a suitable volume of the appropriate growth media to cells, dropwise. Supernatant was removed following centrifugation ( $300 \times g$  for 5 min) and cell pellets were then re-suspended in a suitable volume of the appropriate growth media.

### 2.2.1.4 Retroviral Transduction

$1 \times 10^6$  cells were suspended in 2 mL of media (McCoy's or IMDM, 10% v/v FCS, 2 mM L-Glutamine) containing a recombinant retrovirus that was grown using the  $\Psi$ CRIP cell line (gifted by Richard Mulligan, Whitehead Institute, Cambridge). The retrovirus carried the pBABE-Puromycin vector (Addgene) which contained the gene encoding a dominant-negative human-telomerase reverse transcriptase (TERT) protein and the puromycin-resistance transgene. Cells were incubated at  $37^\circ\text{C}$  with 5%  $\text{CO}_2$  for 4 h, after which 2 mL of supplementary media were added. 24 h after transfection, cells that had successfully integrated the pBABE-Puro vector into their genome were selected for using puromycin and then maintained under constant selection ( $1 \mu\text{g}/\text{mL}$ ) for the duration of the experiment. The puromycin concentration required was tested for in a separate kill-curve experiment. Cells were plated in a 6-well plate and after 24 h refed with different amount of puromycin ( $0.5$  –  $2.5 \mu\text{g}$ ). Cell

death was monitored daily over 5 days. Total cell death in 2-3 days was observed with 1  $\mu\text{g}/\text{mL}$  puromycin.

## **2.3 NUCLEOFECTATION ASSAY**

### **2.3.1 Plasmid extraction**

#### **2.3.1.1 Media and antibiotics**

10 LB-broth tablets (Sigma) were added to 500 mL double-distilled H<sub>2</sub>O (ddH<sub>2</sub>O) and autoclaved (for 2x 200 mL cultures of plasmids). 10 LB-agar tablets (Sigma) were dissolved in 500 mL ddH<sub>2</sub>O and autoclaved. The next day, 500 mL of LB-broth agar was boiled in the microwave at a medium heat for 7 min and then cooled to ~55°C in a water bath. 160 µL of Ampicillin (100 mg/mL, Sigma) was added to 160 mL LB-agar and left > 1 h to set at room temp.

#### **2.3.1.2 Plate streaking and starter cultures**

Ampicillin agar plates were warmed to 37°C before being streaked with ~10 µL each of TALEN L2 or R2 glycerol stock using a sterile spreader. The plates were then inverted and left overnight. Single colonies (L2 and R2 separately) were picked and added to a 50 mL tube with 5 mL LB-broth and 5 µL Ampicillin (Sigma). Tubes were shaken at 220 RPM for 8 h at 37°C (Innova 44 shaking incubator, New Brunswick)

#### **2.3.1.3 Endotoxin-free MaxiPrep**

After 8 h shaking incubation, 150 µL starter culture was inoculated with 150 mL LB-broth containing Ampicillin, overnight. Cultures were pelleted in 3 x 50 mL tubes for 20 min at 2800 x g at 4°C. The Nucleobond Xtra Midi Plus Kit (Macherey-Nagel) was used as per the manufacturer's guidelines to extract and purify the plasmids.

#### **2.3.1.4 Verification**

Following DNA quantification, plasmid sizes were verified by linearizing 500 ng DNA with *Xba*I. 500 ng DNA was added to 2 µL Buffer M, 0.2 µL BSA, 0.3 µL *Xba*I (5 U)

and 17  $\mu\text{L}$  ddH<sub>2</sub>O and incubated at 37°C for 1 h. *Xba*I-cut and uncut DNA samples were resolved using 0.5% agarose gel electrophoresis to analyse size.

### 2.3.2 Transient transfection

Specific telomeric double-strand DNA breaks were induced by transient transfection with TALEN (Transcription Activator-Like Effector Nuclease) plasmid pairs (Liddiard et al. 2016 ; Liddiard et al. 2018). A 17p TALEN pair was used to induce a DSB in the 17p chromosome subtelomere, 14 bp from the start of the telomere and a 16p/21q TALEN pair to induce a DSB 1.5 kb from the start of the telomere in the 16p/21q family of telomeres (Liddiard et al. 2016).

2.5  $\mu\text{g}$  of each TALEN plasmid was used for nucleofection of subconfluent HAP1, RPE1<sup>hTERT</sup> and HCT116 cells. Cells were nucleofected in batches ( $1 \times 10^6$  cells/cuvette in 100  $\mu\text{L}$  supplemented nucleofection mix, SE Cell Line Kit, Lonza) using an Amaxa 4D Nucleofector.

For the HAP1 cell line, the nucleofector program FF-113 was optimised to achieve ~70% transfection efficiency with the pmaxGFP expression vector 24 h post-nucleofection.

For RPE1<sup>hTERT</sup> and HCT116 cell lines, the DS-138 nucleofector program was optimised to achieve 70.3% / 72.3% (RPE1<sup>hTERT</sup> WT/*POLQ*<sup>-/-</sup>) and 80% / 77.6% (HCT116 WT / *POLQ*<sup>-/-</sup>) transfection efficiency with the pmaxGFP expression vector (Lonza) 24 or 48 h post-nucleofection.

Three replicates of each transfection were performed and cultured in individual wells prior to cell harvests at 24 or 48 h post-nucleofection. Replicate transfectants were pooled for counting before pelleting for DNA extraction and samples were taken to perform the Hoechst 33342 and propidium iodide assay (Chemometec Nucleocounter NC-3000) to determine GFP expression as an indicator of transfection efficiency.

## 2.4 DNA/PROTEIN EXTRACTION

### 2.4.1 DNA Extraction from cells

Genomic DNA was extracted by standard Proteinase K, RNase A, phenol/chloroform extraction (Sambrook et al. 1989). Cells ( $\geq 10^5$ ) were lysed with 30  $\mu\text{g}$  RNase A (10 mg/mL, Sigma), 60  $\mu\text{g}$  proteinase K (20 mg/mL, Sigma) 300  $\mu\text{L}$  lysis buffer (10mM Tris-HCl pH 8, 100 mM NaCl, 0.5%, Sodium Dodecyl Sulphate (SDS) and 5 mM EDTA pH 8) for 15 h at 45°C. Following cell lysis, 300  $\mu\text{L}$  of phenol/chloroform/isoamyl alcohol (25:24:1, pH 8, Sigma) was added to the cell lysate and rotated using a tube rotator at room temperature for 30 min.

To separate the phases that were formed following tube rotation, the samples were centrifuged for 5 min at 16,000 x *g*. Both the aqueous phase and the interphase were transferred to a 1.5 mL tube which contained 300  $\mu\text{L}$  phenol/chloroform/isoamyl alcohol. Samples now containing just the aqueous phase and the interphase were rotated again for 20 min at room temperature using a tube rotator. Following centrifugation at 16,000 x *g* for 5 min, only the aqueous phase which contained the DNA was transferred to a 1.5 mL tube. To precipitate the DNA, 900  $\mu\text{L}$  of 100% ice-cold ethanol (-20°C) and 30  $\mu\text{L}$  (1/10 volume) of sodium acetate (3 M, pH 5.3, Sigma) were added to the DNA. For optimal precipitation, samples were stored at -20°C for 15 h. Following centrifugation (5 min 16,000 x *g*), the DNA pellet was washed with 70% ice-cold ethanol. The residual ethanol was completely evaporated by air drying for at least 30 min, following which, the DNA was resuspended in Tris-HCl buffer (10 mM, pH 8).

### 2.4.2 DNA Quantification

Hoechst 33258 fluorometry was used to quantify DNA with a QuantiFluor-ST Fluorometer (Promega). 1xTEN solution (10 x stock; 100 mM Tris, 10 mM EDTA, 2 M NaCl, pH 7.4, prepared with ddH<sub>2</sub>O) was prepared and Hoechst 33258 DNA

intercalator fluorescent dye (1 mg/mL stock, Bio-Rad) was then added to a final concentration of 0.1  $\mu\text{g/mL}$ . To calibrate the fluorometer, 2 mL of 1x TEN solution (containing Hoechst) was read (blank). For the 500 ng standard, 10  $\mu\text{L}$  calf thymus DNA (100  $\mu\text{g/mL}$ ) was mixed with 2 mL of 1xTEN solution containing Hoechst. For sample DNA, 2  $\mu\text{L}$  of DNA were added to 2 mL of 1x TEN assay buffer and mixed thoroughly before being quantified according to the manufacturer's instructions. Samples were quantified in triplicate.

## 2.5 POLYMERASE CHAIN REACTION (PCR)

The Polymerase Chain Reaction (PCR) technique was used to amplify specific DNA fragments.

### 2.5.1 Conventional PCR

Typically, 50–200 ng of DNA per  $\mu\text{L}$  were used for each reaction and the reagents listed in **Table 2.1** were added. 10x Taq buffer used consisted of 75mM Tris-HCl (pH 8.8), 20 mM  $(\text{NH}_4)\text{SO}_4$  and 0.01% Tween-20. A 10:1 mixture of *Taq:Pwo* polymerase (500 U/100  $\mu\text{L}$ , Thermo Scientific) was used unless stated otherwise.

**Table 2.1 Conventional PCR reagents**

Reagent (Stock)	1x reaction ( $\mu\text{l}$ )
ddMiliQ $\text{H}_2\text{O}$	14.05
Taq Buffer (10x)	2.5
$\text{MgCl}_2$ (25 mM)	2
dNTPs (100 nM)	0.3
Forward primer (10 $\mu\text{M}$ )	2.5
Reverse primer (10 $\mu\text{M}$ )	2.5
Genomic DNA (50 – 200 ng/ $\mu\text{L}$ )	1
<i>Taq</i> Polymerase (500 U/100 $\mu\text{L}$ )	0.25
<b>Total</b>	25 $\mu\text{L}$

The reactions were cycled using a Tetrad thermal cycler (Bio-Rad) under the following conditions: 94°C for 20 s, 52-68°C for 30 s followed by 68°C for 1 min/kb. A gradient (52-68°C) was used for optimisation of the primer annealing temperature when required.

### 2.5.2 STELA Polymerase Chain Reaction (PCR)

Individual telomeres from specific chromosome ends were amplified using STELA PCR (Baird et al. 2003) (Genomic DNA samples were diluted to a final concentration of 250 pg/ $\mu\text{L}$  and mixed with 250 pM Telorette2 (Tel2) linker to a volume of 40  $\mu\text{L}$  of

Tris HCl (10 mM, pH 8.0). 1  $\mu$ L of this DNA/Tel2 mixture was added to a 10  $\mu$ L reaction containing the components listed in **Table 2.2**. For each sample, 4-6 replica PCR reactions were performed using the same mixture of reagents.

**Table 2.2** STELA PCR reaction reagents

Reagent (Stock)	1x reaction ( $\mu$ l)
ddMiliQ H <sub>2</sub> O	4.98
Taq Buffer (10x)	1
MgCl <sub>2</sub> (25 mM)	0.8
dNTPs (100 nM)	0.12
Telomere specific primer (5 $\mu$ M)	1
Teltail primer (5 $\mu$ M)	1
DNA/Tel2 mix	1
<i>Taq/Pwo</i> (10:1U)	0.1
<b>Total</b>	10 $\mu$ l

The annealing temperature used was dependent on the telomere-specific primer used. Reactions were cycled using a Tetrad thermal cycler (Bio-Rad) using the following conditions for STELA using the XpYpE2 primer: 22 cycles of 94°C for 20 s, 65°C for 30 s, 68°C for 8 min. For each experiment, a positive PCR control (DNA previously shown to produce a robust STELA profile) was included. Additionally, STELA products from a previous experiment that was known to work was run on the gel as a hybridisation control.

### 2.5.3 Fusion PCR

Telomere fusion amplicons were amplified using single-molecule telomere fusion PCR (Capper et al. 2007; Letsolo et al. 2010). This multiplex PCR involved the use of primers that could detect 24 specific chromosome ends. Of the 5 primers available, 3 target specific chromosomes (XpYp, 5p and 17p) whereas 2 target subtelomeric families (16p and 21q):

*16p family*: 16p, 1p, 9p, 12p, 15q, XqYq and 2q14 interstitial telomeric locus.

*21q family*: 21q, 1q, 2q, 5q, 6q, 6p, 8p, 10q, 13q, 17q, 19p, 19q, 22q and the 2q13 interstitial telomeric locus (Capper et al. 2007).

DNA was diluted to 50-200 ng/ $\mu$ L in 10 mM Tri-HCl (pH 8.0). Multiple (4-12) 10  $\mu$ L reactions were set up using the reagents listed in **Table 2.3**.

**Table 2.3** Example: 16p6:21q1 fusion PCR reaction mix

Reagent (Stock)	1x reaction ( $\mu$ l)
ddMiliQ H <sub>2</sub> O	3.98
Taq Buffer (10x)	1
MgCl <sub>2</sub> (25 mM)	0.8
dNTPs (100 nM)	0.12
16p6 (10 $\mu$ M)	0.5
21q1 (10 $\mu$ M)	0.5
DNA (50-200 ng/ $\mu$ L)	1
Taq/Pwo (10:1U)	0.1
<b>Total</b>	10 $\mu$ l

Reactions were then cycled under the following conditions: 25 cycles of 94°C for 20 s, 62°C for 30 s and 68°C for 8 min using a Tetrad thermal cycler (Bio-Rad).

## **2.6 GEL ELECTROPHORESIS**

### **2.6.1 STELA and Fusion amplicons**

STELA and fusion PCR amplicons were resolved using low percentage 0.5% agarose Tris-acetate-EDTA (TAE) gel electrophoresis. The agarose (Roche) was dissolved in boiling Tris-acetate-EDTA (40 mM Tris base, 20 mM acetic acid, 1 mM EDTA) and then cooled to 50°C before adding 40 µL Ethidium Bromide (EtBr) (0.625 mg / mL). The gel was submerged in 1x TAE buffer that was cooled to 4°C using a circulating cooling system (Grant) to enhance resolution. 5 µL of the STELA or fusion PCR products were mixed with 1 µL of a 6x loading dye (5% bromophenol blue, 5% xylene, 15% Ficoll). PCR products were loaded in pre-set individual wells and run at 120 volts (V) for 16 h on a 40 cm gel. Fusion PCR products were run at ~200 V for ~2 h or for 50 V for 16 h. As a 10 cm gel was sufficient for the suitable resolution of fusion PCR products, a 40 cm gel was used in which a second and third row of pre-set wells could be set up. DNA ladders were visualised using a UV-transilluminator (Flowgen).

### **2.6.2 Standard DNA products**

PCR products were resolved using agarose (0.7-1%; Geneflow) Tris-acetate-EDTA gel electrophoresis, with 10-20 cm gels at 100 V for ~2 h. The percentage of the agarose gel was determined based on the expected size of the DNA fragment produced. DNA bands were visualised using a UV-transilluminator (EPS).

## 2.7 SOUTHERN BLOTTING AND RADIOLABELLING

### 2.7.1 Southern Blotting

Agarose gels containing STELA or fusion products, were washed twice in depurination buffer (0.25 M HCl) for 6 min/wash to depurinate the DNA, thereby fragmenting it into smaller fragments which facilitates the later transfer to the nylon membrane. Following the removal of the depurination buffer, the gel was rinsed in distilled H<sub>2</sub>O (dH<sub>2</sub>O) before being washed in denaturation buffer (1.5 M NaCl, 0.5 M NaOH) for 15 min to denature the dsDNA and enable hybridisation. Southern blotting was used to transfer DNA to a positively charged nylon membrane (Hybond-XL, GE Healthcare) by placing the membrane on top of the agarose gel. A stack of paper towels and a weight were placed on top of the membrane to ensure DNA fragments were transferred by alkaline Southern blotting (>3 h) using denaturation buffer.

### 2.7.2 Radiolabelling DNA probe

DNA probes that were specific for the telomere repeat sequences were used to detect STELA products. Likewise, pairs of telomere-adjacent primers that were specific for each of the chromosomes ends involved in the fusion assay were generated by PCR. DNA probes were radiolabelled using random oligonucleotide labelling (Feinberg and Vogelstein 1983). A solution of 45 µL TE buffer (10 mM Tris-HCl and 1 mM EDTA) that contained 25 ng of a DNA probe and 5 µL of Megaprime primer (GE healthcare) was heated to 96°C for 5 min to denature the DNA. The solution was cooled on ice for 5 min, following which 10 µL Megaprime labelling buffer, 1 µL megaprime enzyme and 4 µL [<sup>33</sup>P] dCTP (3000Ci/mmol, PerkinElmer) was added. The solution was then incubated in a 37°C water bath for 1 h. After this, 1 µL of a [<sup>33</sup>P] dCTP-labelled DNA ladder (1 kb and 2.5 kb) and 50 µL ddH<sub>2</sub>O water was added. Radiolabelled probes were stored at 4°C for up to 2 weeks.

### **2.7.3 Hybridisation**

Hybond membranes (GE Healthcare) were rinsed in water following Southern blotting before being incubated at 56°C in pre-warmed hybridisation bottles with 15 mL of pre-warmed Church buffer (7% SDS, 0.5 M sodium phosphate buffer (1 M disodium hydrogen phosphate and 1 M sodium dihydrogen phosphate), 1 mM EDTA, 1% BSA, pH 7.2). The radiolabelled probe was denatured by heating to 96°C for 5 min. 25 µL of denatured DNA probe (for TRF blots, 50 µL telomere probe was used) was added to the hybridisation bottle and left to hybridise in a rotating hybridisation oven at 56°C for a minimum of 4 h. Excess probe was removed by washing the Hybond membrane in wash buffer (0.1% saline sodium citrate and 0.1% sodium dodecyl sulphate) at 56°C for 15 min. This was repeated three times. STELA blots were dried at 60°C for 30 min, whereas excess water was removed from fusion blots before they were wrapped in cling film to ensure that blots did not dry out completely.

### **2.7.4 Visualisation of radiolabelled membranes**

Radiolabelled membranes were placed into an exposure cassette for 15 h with a molecular dynamic phosphorimaging screen (Amersham) placed on top. The screen was scanned using a Typhoon FLA 9500 biomolecular imager (GE Healthcare) to detect hybridised fragments. Radiolabelled blots were visualised using ImageQuant TL (GE Healthcare).

### **2.7.5 Stripping and re-probing membranes**

Radiolabelled probes were stripped from fusion blots by washing the membrane in a boiling solution of 500 mL of 0.2% SDS for 1 h, repeated three times. To ensure stripping was complete, the membranes were visualised again (as previously described) to check that there was no residual signal. Membranes were then hybridised with the next probe. Probes that give the lowest fusion frequency were

used first so that stripping between probes was easier and re-hybridised membranes were cleaner. Thus, sequential hybridisations in the order 17p – 16p – 21q were undertaken with each of the telomere-specific probes.

### **2.7.6 Gel Analysis**

To calculate the molecular weights of each STELA telomere amplicon, ImageQuant TL (GE Healthcare) was used to compare the resolved telomeres with the DNA ladder used. To account for the actual telomere size, the distance between the binding site of the telomere adjacent primer to the start of the telomere was subtracted (E.g. XpYpE2 is 408 bp from the start of the telomere). Statistical analysis was performed using GraphPad Prism 5.

## 2.8 TERMINAL RESTRICTION FRAGMENT (TRF) ASSAY

The TRF assay (Kimura et al. 2010) was used to detect the terminal restriction fragment length of chromosomal DNA containing the telomere and undigested subtelomeric DNA. More information about this technique can be found in *Chapter 1*. Genomic DNA (1.5 µg) was restriction digested with 30 U *Hinf1* and 30 U in *Rsa1* in 1x Buffer M (New England Biolabs) at 37°C for >3 h. To determine whether the DNA was fully digested, an aliquot of the reaction was resolved with 0.7% agarose gel electrophoresis to compare 500 ng of undigested and digested samples. The DNA was visualised following ethidium bromide staining. Successfully digested samples were mixed with 6x loading dye (5% bromophenol blue, 5% xylene, 15% Ficoll) and resolved using 0.5% agarose gel electrophoresis. Samples were run at 100 V for ~ 1 h followed by Southern blotting and radiolabelling (*Section 2.7.1-4*) to detect telomere length. Telomere length was calculated using ImageQuant by comparing telomere length profiles with standard DNA ladders that were also run on the gel.

## 2.9 TELOMERASE ASSAY

### 2.9.1 Protein Extraction

Approximately  $1 \times 10^6$  cells were centrifuged for 5 min at  $300 \times g$  following suspension in PBS. The supernatant was removed, and the cell pellet was suspended in 200  $\mu\text{L}$  of CHAPS lysis buffer (1x, Millipore) and incubated for 30 min on ice. Cell lysates were spun at  $12,000 \times g$  ( $4^\circ\text{C}$ ) for 20 min. The protein-containing supernatant was then aliquoted for protein quantification and for use in the TRAP assay.

### 2.9.2 Protein Quantification

Protein quantification was performed using the Coomassie Plus Assay Kit (Thermo Scientific) and BSA (2 mg/mL) as a standard. 10  $\mu\text{L}$  of each standard or supernatant sample were combined with 300  $\mu\text{L}$  of Coomassie Plus Protein Assay Reagent and allowed to incubate at room temperature for 10 min. Absorbance at 595 nm was then measured using the FLUOstar Optima Microplate Reader (BMG Labtech). Standards and supernatant samples were measured in duplicate. Protein concentration was extrapolated from the BSA standard curve readings.

### 2.9.3 Telomerase Repeated Amplification Protocol (TRAP)

In order to detect telomerase activity, the Telomerase Repeated Amplification Protocol (TRAP) was followed (Mender and Shay 2015). Reactions were prepared using 5  $\mu\text{L}$  protein samples (100 ng/ $\mu\text{L}$ ). Reactions were set up with the primer mix (**Table 2.4**), 2x TRAP buffer (**Table 2.5**) and reagents listed in **Table 2.6**. DEPC  $\text{H}_2\text{O}$  was required to ensure that RNase enzymes were inactivated and did not affect the reactions. Reactions (45  $\mu\text{L}$  master mix with 5  $\mu\text{L}$  protein sample) were then cycled under the following conditions:  $25^\circ\text{C}$  for 40 min,  $95^\circ\text{C}$  for 5 min then 29 cycles of  $95^\circ\text{C}$  for 5 s,  $52^\circ\text{C}$  for 30 s,  $72^\circ\text{C}$  for 45 s followed by  $72^\circ\text{C}$  for 10 min using a SimpliAmp thermal cycler (Thermo-Fisher).

**Table 2.4 TRAP primer mix reagents**

Primer mix reagents (stock)	Volume ( $\mu\text{L}$ )
ACX (100 ng/ $\mu\text{L}$ )	2.22
NT (100 ng/ $\mu\text{L}$ )	3.7
TSNT (5x)	1
DEPC H <sub>2</sub> O	13.08
<b>Total</b>	20

**Table 2.5 2x TRAP buffer reagents**

TRAP buffer reagents (stock)	Volume
Tris-HCl (pH 8.0)	400 $\mu\text{L}$
MgCl <sub>2</sub> (25 mM)	1.2 mL
KCl (2 M)	633 $\mu\text{L}$
Tween 20 (100%)	1 $\mu\text{L}$
EGTA (0.5 M)	250 $\mu\text{L}$
BSA (2mg / mL)	1 mL
dCTP (100 mM)	10 $\mu\text{L}$
dATP (100 mM)	10 $\mu\text{L}$
dGTP (100 mM)	10 $\mu\text{L}$
dTTP (100 mM)	10 $\mu\text{L}$
DEPC H <sub>2</sub> O	6.48 mL
<b>Total</b>	10 mL

**Table 2.6 TRAP master mix reagents**

TRAP master mix reagents (stock)	1 reaction ( $\mu\text{L}$ )
DEPC H <sub>2</sub> O	17.6
TRAP buffer (2x)	25
TS primer (100 ng/ $\mu\text{L}$ )	1
Primer mix	1
Titanium <i>Taq</i> polymerase (hot-start <i>Taq</i> )	0.4
<b>Total</b>	45

### 2.9.4 Electrophoresis

TRAP assay products were resolved using 5xTBE (Tris base, H<sub>3</sub>BO<sub>3</sub>, EDTA, dH<sub>2</sub>O) 40% acrylamide (19:1) gel electrophoresis (**Table 2.7**). TRAP products were mixed with 10 µL of a 6x loading dye (5% bromophenol blue, 5% xylene, 15% Ficoll) per sample and then loaded into 10 µL pre-set individual wells and run at 100 V for ~1.5 h.

**Table 2.7** TBE gel components

Gel component (stock)	Volume
Distilled H <sub>2</sub> O	16.25 mL
40% acrylamide (19:1)	6.25 mL
TBE (5x)	2.5 mL
APS (10%)	125 µL
TEMED	25 µL
<b>Total</b>	20 mL

### 2.9.5 Visualisation of TRAP products

TBE gels containing resolved TRAP products were stained with SYBR gold (1:10000, Thermo-Fisher) by gentle rotation for 10 min at room temperature. Stained gels were then visualised using a Typhoon FLA 9500 biomolecular imager (GE Healthcare).

## 2.10 PROTEIN IDENTIFICATION

I performed western blotting with two commercially available anti-POLQ antibodies. Professor Richard Wood's laboratory (MD Anderson Cancer Center, University of Texas) also performed western blotting and immunoprecipitation.

### 2.10.1 Nuclear extracts

Professor Richard Wood's laboratory generated nuclear extracts of HAP1, HCT116 and RPE1<sup>hTERT</sup> cell lines using a nuclear extract protocol from Rockland Antibodies & Assays (<https://rockland-inc.com/Nuclear-Extract-Protocol.aspx>) (Baldwin 1996). These experiments were performed by Megan Lowery (BSc, University of Texas).

### 2.10.2 Western Blotting

#### 2.10.2.1 Protein Extraction

Cells were washed twice in PBS and then centrifuged at 100 x *g* (4°C) for 5 min. Supernatant was discarded and the cells were suspended in 2.5x their volume of lysis buffer (150 mM NaCl, 50 mM Tris-HCl, 5 mM EDTA, 1% NP40) that also contained 3% v/v PMSF (100 mM, Sigma), 1% v/v Protease Inhibitor Cocktail III (100 mM AEBSF Hydrochloride, 80 μM Bovine Lung Aprotinin, 5 mM Bestatin, 1.5 mM E-64 Protease Inhibitor, 2 mM Leupeptin Hemisulfate, 1 mM Pepstatin A, Calbiochem) and 1% v/v Phosphatase Inhibitor Cocktail II (200 mM Imidazole, 100 mM Sodium Fluoride, 115 mM Sodium Molybdate, 100 mM Sodium Orthovanadate, 400 mM Sodium Tartrate Dihydrate, Calbiochem). Cells were left on ice for 5 min before being centrifuged at 20,000 x *g* for 30 min (4°C). The protein-containing supernatant was removed and stored at -80°C until required.

### **2.10.2.2 Protein Quantification**

Protein samples were compared with a standard curve prepared from serially diluted purified Bovine Serum Albumin (BSA) protein. Therefore, (2 mg/mL) was diluted to 0, 100, 200, 300 and 400 µg/mL using lysis buffer. The protein samples to be quantified were diluted 1/10 in ddH<sub>2</sub>O. 10 µL of either a BSA standard or a protein sample were added to a 96-well plate and combined with 300 µL of Coomassie blue reagent (Thermo Scientific) and allowed to incubate at room temperature for 10 min. The FLUOstar Optima Microplate Reader (BMG Labtech) was used to measure absorbance at 595 nm. Protein sample concentrations were extrapolated from the BSA standard curve.

### **2.10.2.3 Electrophoresis**

5 µL of 3x loading buffer (0.5M Tris-HCl, 30% Glycerol, 10% SDS, 3% 2-mercaptoethanol, 1% bromophenol blue) was added to 5–20 µg of protein in a volume of 15 µL. Protein samples and molecular weight markers were then boiled at 100°C for 5 min. Proteins were resolved using a 4-20% Mini-Protean TGX Precast Gel (Bio-Rad), submerged in running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3), and run at 100 volts for 1-2 h.

Professor Richard Wood's laboratory used 3–8% Criterion™ XT Tris-Acetate Protein 26 well Gel (15 µL well capacity). Precision Plus Protein All Blue Standards (Bio-rad) and protein samples were heated to 65°C for 3 min. Gels were submerged in with 1x running buffer (250 mM Tris-HCl, pH 6.8, 50% glycerol, 10% SDS, 0.2% bromophenol blue and 500 mM DDT) at 75 V for 240 min.

### **2.10.2.4 Transfer**

A polyvinylidene fluoride transfer membrane (PVDF) (Immobilon-P, Millipore) was submerged in methanol (Fisher), before being washed in dH<sub>2</sub>O for 2 min by gentle

rotation and then transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, pH 8.3) for 1 min by gentle rotation. Resolved proteins were transferred from the 4-20% Mini-Protean TGX Precast Gel (Bio-Rad) to the PVDF membrane by western blotting in transfer buffer for 2 h at 100 volts (4°C).

Data received from Professor Richard Wood's lab was produced using both a PVDF (Immobilon-P, 0.45 µm, Millipore) and nitrocellulose membranes for protein transfer.

#### **2.10.2.5 Immunoprobng**

Membranes were blocked in 0.2% Tween20/PBS/5% milk for 1 h, following which, membranes were placed into a 50 mL tube containing diluent (50 mL Tween20/PBS, 5% milk and 1% BSA) and the appropriate concentration of primary antibody (e.g. 1/1000 POLQ antibody). 50 mL tubes were left on a roller at 4°C overnight. Membranes were washed using 1% Tween20/PBS for 25 min (1% Tween20/PBS was changed every 5 min). The membranes were added to a new 50 mL tube which contained the appropriate antibody solution and the appropriate concentration of secondary antibody (**Table 2.8**). 50 mL tubes were left on a roller at room temperature for 1 h following which membranes were washed using 1% Tween/PBS for 25 min (1% Tween/PBS was changed every 5 min).

**Table 2.8** A list of the primary and secondary antibodies, the concentrations and the immunoprobings solutions used. All secondary antibodies were conjugated to horseradish peroxidase (HRP).

Primary Antibody (Dilution)	Primary Antibody (Solution)	Secondary Antibody (Dilution)	Secondary Antibody (Solution)
Mouse Monoclonal Anti-POLQ (2 $\mu$ L in 2 mL) Abnova (M09), Clone 1C11	0.2% Tween20/ PBS/5% Milk	Anti-Mouse HRP (1 $\mu$ L in 2 mL)	0.2% Tween20/ PBS/5% Milk
Rabbit Polyclonal Anti-POLQ (1 $\mu$ L in 2 mL) Aviva Systems Biology ARP49203_P050-HRP	0.2% Tween20/ PBS/5% Milk	Anti-Rabbit HRP (1 $\mu$ L in 2 mL)	0.2% Tween20/ PBS/5% Milk
Rabbit Polyclonal Anti-Actin Sigma A2066 (1.3 $\mu$ L in 1 mL)	0.2% Tween20/ 1% BSA/5% Milk/PBS	Anti-Rabbit HRP (1 $\mu$ L in 2 mL)	0.2% Tween20/ 1% BSA/5% Milk/PBS
Mouse Monoclonal Anti-POLQ Antibody 5 (Atb5,153-5-1) Professor Wood's laboratory (2.39 mg / mL)	10X Tris Buffered Saline/ Tween 20/ 10% milk	Anti-mouse HRP (1 in 10,000)	20 mM Tris, pH 7.5/137 mM NaCl / 0.1% Tween20/ 5% Milk

### 2.10.2.6 Visualisation with luminescent substrates

Luminescent detection was used as it can be completed quickly, and signal is detected by film which is easier to photograph and quantitate by densitometry compared with chromogenic development. Furthermore, varying the exposure can result in increased sensitivity. The ECL Plus kit (Thermo Scientific) was used for visualisation of the membrane. In a dark room, Chemiluminescence Hyperfilm ECL (GE Healthcare Amersham) was placed on top of the membrane before being developed and fixed using the SRX-101A Medical Film Processor (Konica Minolta Medical and Graphic Inc.). Molecular weight markers (Precision Plus Protein Dual Colour Standards, Biorad) of known sizes were used as a comparison to identify proteins by their molecular weight.

**2.10.2.7 Stripping antibody from membranes**

Membranes were washed for 5 min in dH<sub>2</sub>O, 5 min in NaOH (0.2 M) and a further 5 min in dH<sub>2</sub>O by gentle rotation. Following washing, membranes could undergo immunoprobng with a different antibody (2 mL diluent and 1/2000 mouse secondary antibody).

## **2.11 SANGER SEQUENCING**

### **2.11.1 DNA extraction from agarose gel**

Sanger sequencing was used to confirm CRISPR/Cas9-induced mutations in *POLQ* for parental cell lines used. To prepare DNA samples for Sanger sequencing, agarose gels were used to electrophorese PCR products and then the DNA was excised using a clean scalpel. DNA was extracted from the agarose using the Monarch DNA Gel Extraction kit (New England Biolabs). DNA was eluted in  $\leq 20$   $\mu\text{L}$  ddH<sub>2</sub>O and any unused sample was stored at 4°C for subsequent use.

### **2.11.2 Sanger Sequencing using Eurofins**

Premixed samples were prepared by mixing 2  $\mu\text{L}$  of primer (10  $\mu\text{M}$ ) with 15  $\mu\text{L}$  of the purified PCR product (10 ng/ $\mu\text{L}$  for 300-1000 bp). These samples were then sent to Eurofins Genomics for sequencing. Sequenced DNA received from Eurofins Genomics was analysed using Basic Local Alignment Search Tool (BLAST) software on the NCBI (<https://www.ncbi.nlm.nih.gov/>) and Ensembl ([www.ensembl.org](http://www.ensembl.org)) websites.

## **2.12 HIGH THROUGHPUT SEQUENCING OF TELOMERE FUSION PRODUCTS**

### **2.12.1 PCR clean-up**

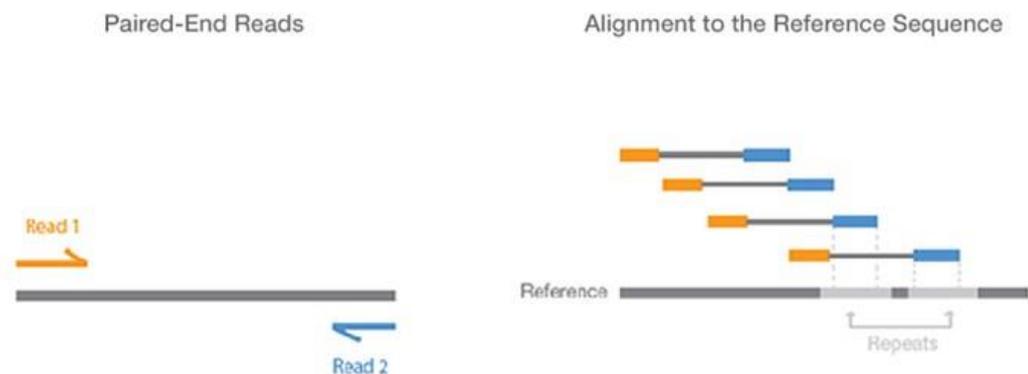
PCR reagents (primers, nucleotides, enzymes and salts) were removed from PCR products by magnetic separation using the Agencourt AMPure XP purification kit (Beckman Coulter) and the DynaMag-2 magnet (Life Technologies).

For each sample used to generate fusion amplicons, 96 fusion PCR reactions were prepared, and 5 random reactions were sampled for quality control validation by Southern blotting. The remaining reactions were pooled and purified. A 5  $\mu$ L aliquot was removed to compare products before and after purification to ensure there was no loss or altered profile of purified products. The remaining reactions were mixed with AMPure XP reagent (1.8 mL AMPure XP/1 mL PCR product) at room temperature for 5 min, following which, PCR products were bound to the paramagnetic beads. Tubes containing the reactions/AMPure XP reagent were then transferred to the DynaMag-2 magnet for 2 min to allow separation of the beads bound to the PCR products from the PCR reagent contaminants. The supernatant containing the contaminants was discarded.

The PCR products bound to the paramagnetic beads were washed with 70% ethanol twice to remove any remaining contaminants. The ethanol was then discarded, and the paramagnetic beads were left to air-dry for 20 min to remove residual ethanol. Tubes were then removed from the DynaMag-2 magnet and the PCR products eluted from the beads with nuclease-free H<sub>2</sub>O. Tubes were transferred to the DynaMag-2 magnet again for 2 min to allow the beads to separate from the eluted DNA sample. The purified PCR products were then pooled and a 5  $\mu$ L aliquot was used to compare PCR products before and after purification by using 0.5% agarose gel electrophoresis followed by Southern blotting and hybridisation of the membrane with a radiolabelled fusion probe.

### 2.12.2 Illumina HiSeq4000 paired-end High Throughput Sequencing

Illumina HiSeq4000 paired-end sequencing of 5 HAP1 WT<sup>DN-hTERT</sup> and 5 HAP1 POLQ<sup>:DN-hTERT</sup> clones going through telomere-driven crisis was performed by the Wales Gene Park. Paired-end sequencing was chosen to sequence mixed pools of fusion amplicons generated from each sample. As the distance between each paired-read is known, telomere fusions can be detected (**Figure 2.1**). Moreover, repetitive regions (such as telomeric DNA) can be sequenced with high mapping quality using this method (Liu et al. 2012).



**Figure 2.1** Taken from ([www.illumina.com](http://www.illumina.com)). Paired-end high-throughput sequencing enables both ends of the DNA fragment to be sequenced. Alignment algorithms use the distance between paired reads to map the reads over repetitive regions more precisely.

### 2.12.3 Fusion calling using Dysgu

To identify fusion events, a novel analysis software named Dysgu, (available online at <https://github.com/kcleal/dysgu>) was developed by Dr Kez Cleal (Cardiff University). This software takes aligned sequencing reads in BAM format as input, and outputs a table of putative fusion events with each fusion event assigned a probability value. Dysgu implements several innovations, including ‘directed mapping’ which selectively maps subtelomere derived reads to the expected genome region, whilst also being agnostic of reads derived from other loci. Dysgu also incorporates a novel read-pairing algorithm based on the concept of an optimal ‘path’ of alignments

(Cleal et al. 2019). Dr Kez Cleal (Cardiff University) performed the data handling, quality control (QC) and mapping strategies. I performed the downstream analysis such as the manual curation and investigation of telomere fusion events.

#### **2.12.4 Manual curation and downstream analysis**

Telomere fusion events called by Dysgu were classified into distinct type of events. Fusions between 16p/21q family subtelomeres and genomic loci were classified as 'genomic' telomere fusion events. Fusion events between 16p and 21q family subtelomeres were categorised as 16p-21q inter-chromosomal fusion events. Any fusion events between 16p or 21q family subtelomeres (16p-16p or 21q-21q) were classified as '16p' or '21q' events. Fusion events were visualised using the tool Integrative Genome Viewer (IGV) (Robinson et al. 2011). 100 fusion events of each type were manually validated for each clone by comparing fusion sequences against the human genome and subtelomeric sequences using BLAST. This included mapping microhomology and insertions identified at the fusion junction. The genomic loci incorporated into telomere fusion events were also studied. More information can be found in *Chapter 5*.

### **Calculating predicted fusion frequencies**

Fusion frequency was adjusted to chromosome size to determine whether any chromosomes had a particular abundance or deficiency for fusion events. The size of each chromosome was obtained from the Ensembl Sanger website.

The *actual* fusion frequency per chromosome was calculated as follows:

$$\textit{number of fusion junctions per chromosome} / \textit{chromosome size (Mb)}$$

The average fusion frequency for the whole genome was calculated as follows:

$$\textit{total number of fusion junctions} / \textit{whole genome size (Mb)}$$

The *predicted* fusion frequency per chromosome was calculated as follows:

$$\textit{average fusion frequency (whole genome)} \times \textit{single chromosome size (Mb)}$$

### 2.13 WHOLE GENOME SEQUENCING

Genomic DNA samples were derived from 8 HAP1 WT<sup>DN-hTERT</sup> and 8 HAP1 *POLQ*<sup>DN-hTERT</sup> clones both before and after telomere-driven crisis, alongside HAP1 WT and HAP1 *POLQ* parental cell lines. These genomic DNA samples (~1-5 µg per sample) were sent to BGI Tech Solutions (Hong Kong) for whole genome sequencing (WGS). DNA mapping, QC and variant calling were performed by Dr Kez Cleal (Cardiff University, Sections 2.13.1 – 2.13.5). I performed manual curation and downstream analysis.

#### 2.13.1 DNA sequencing and mapping

DNA 100 bp paired-end sequencing reads were generated by BGI using the BGISEQ500 platform. Sequencing was performed to 15x target depth, generating 100 bp read-pairs with the approximate insert size of 250 bp and a standard deviation of 50 bp. Read-pairs were mapped to the GRCh38 reference using bwa mem (v0.7.15), applying the “-M” tag to mark secondary mapping (Li and Durbin 2009). Duplicates were marked using Picard Tools “MarkDuplicates” (v2.7.1), whilst sorting and indexing were performed using samtools (v1.4) (Li et al. 2009).

#### 2.13.2 Copy Number (CN) detection

CN profiles were determined according to Cleal et al. 2019. Briefly, changes in sequencing depth were analyzed in non-overlapping genomic bins of size 10 kb. Bins that had aberrant GC content (< 30 % or > 60 %) or low “mapability” (< 70) defined by the UCSC track ‘wgEncodeCrgMapabilityAlign50mer.bigWig’ (Rosenbloom et al. 2015) were discarded. Also, bins with a coverage value of > 500 were discarded. Bin values were normalized to the genomic median and smoothed with a Haar wavelet transform function using Scipy and PyWavelets (Jones et al. 2001).

Copy number profiles were then segmented using `copyNumber` in R ( $\gamma=4000$ ,  $k_{\min}=10$ ) (Nilsen et al. 2012). Relative copy number changes were identified by using a background subtraction, using either the parental sample as background for before-crisis samples, or using before-crisis or parental profiles for an after-crisis sample. CN heat maps were then generated using the `copyNumber` package.

For CN quantification of gains and losses, segments that corresponded to the genomic background were removed before analysis of copy number variants. Background segments were identified if the segment start and end coordinates occurred in all clonal samples, or if both the segment mean (relative copy number) was in the range -0.6 to 0.6, and the step-change with the subsequent copy number segment was  $< 0.4$ .

### **2.13.3 Structural Variant (SV) calling**

SVs were called using TIDDIT (v2.4.0) using default parameters (Eisfeldt et al. 2017). SVs with both break sites lying on non-canonical reference chromosomes (i.e. chrM etc) were discarded. Translocations with more than 60 supporting reads were discarded, or variants  $< 500$  bp. Variants were then filtered to find unique events, defined as an event that was absent from the parental lines as well all other before-crisis samples. A filtering strategy was followed according to (Cleal et al. 2019). Briefly, each event was cross-checked against parental and other before-crisis samples for uniqueness, searching alignment files for discordant reads that matched the event in question. For translocations and intra-chromosomal variants  $\geq 2.5$  kb, reads were collected from intervals surrounding the break with an interval of 1.25 kb. If a read-pair was found with reciprocal overlap  $> 0.8$ , the call was labelled non-unique. For variants  $< 2.5$  kb the interval was set at 1 kb and the overlap threshold was set at 0.9. A further condition was also set for variants  $< 2.5$  kb in size; only read pairs with the same read orientation to the variant in question were tested.

#### **2.13.4 Insertion and microhomology identification for WGS**

To identify insertions and microhomology, contigs were first assembled from SV derived reads according to Cleal et al. 2019. Briefly, read-pairs supporting SVs were extracted from intervals surrounding breakpoints (pad size 350 bp) before assembly using SPAdes (Bankevich et al. 2012). Contigs were then mapped to GRCh38 using BLAST (Kielbasa et al. 2011) to generate a candidate set of alignments, from which a spanning set of alignments was chosen using fnfi, (available online at <https://github.com/kcleal/fnfi>). Microhomology was identified when two consecutive alignments on the contig overlapped, whereas insertions were identified as gaps between consecutive alignments on the contig (Cleal et al. 2019).

#### **2.13.5 Calculation of gene content of genome**

A BED file of gene intervals that encode transcribed DNA was downloaded from BioMart (available online at <https://www.ensembl.org/biomart>), selecting Ensembl Genes 99 dataset, and the GRCh38 reference. Gene intervals were then merged using bedtools and the total size of all intervals were divided by the length of the reference genome (3,099,734,149 bp <https://www.ncbi.nlm.nih.gov/grc/human/data>).

## Chapter 3

# How does the absence of POLQ impact on the ability of cells to escape a telomere-driven crisis?

### 3.1 ABSTRACT

It has previously been shown that DNA polymerase theta (POLQ) is a crucial component of the alternative non-homologous end-joining pathway (A-NHEJ). This repair mechanism processes DSBs throughout the genome, including DSBs within telomere repeat arrays. Critically short telomeres can become substrates for A-NHEJ during telomere-driven crisis, resulting in the formation of telomere fusion events which may facilitate the escape from crisis. Furthermore, it has been shown that POLQ is overexpressed in several cancer types. In this study I examined the role that POLQ plays in facilitating the escape from a telomere-driven crisis.

I have characterised three human cell lines in which *POLQ* has been disrupted: HAP1, HCT116 and RPE1<sup>hTERT</sup>. The disruption of *POLQ* did not affect growth rates or morphology in any of the cell lines in comparison to the wild type (WT). However, I observed that the *POLQ*-deficient cell lines had longer telomere lengths than the WT for all three cell lines. I performed retroviral transduction with a dominant negative hTERT (DN-hTERT) expression cassette to abrogate telomerase activity in HAP1 WT, HAP1 *POLQ*<sup>-/-</sup>, HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cells and force them into a telomere-driven crisis. Both HAP1 *POLQ*<sup>-/-</sup>:DN-hTERT and HCT116 *POLQ*<sup>-/-</sup>:DN-hTERT clones progressed through and escaped crisis more readily than their WT counterparts. In addition, the onset of a telomere-driven crisis for both HAP1 and HCT116 *POLQ*<sup>-/-</sup>:DN-hTERT clones commenced at longer telomere lengths than the WT<sup>DN-hTERT</sup> clones ( $p < 0.001$  and  $p < 0.01$  respectively, Mann Whitney test). The escape from crisis for both WT<sup>DN-hTERT</sup> and *POLQ*<sup>-/-</sup>:DN-hTERT clonal populations was marked by the stabilisation of telomere length following the reestablishment of telomerase activity.

### 3.2 INTRODUCTION

Telomeres shorten with each round of cell division and telomere length limits the replicative capacity of normal cells. Once telomeres reach a length at which they are no longer able to protect the ends of chromosomes, a TP53 dependent G<sub>1</sub>/S-phase cell cycle arrest is triggered. This is referred to as replicative senescence, during which, although cells are metabolically active, they are unable to divide (Hayflick 1965). Telomere erosion is therefore considered a tumour suppressive mechanism as the growth capacity of cells which have acquired genetic abnormalities over time is limited (Deng et al. 2008). Telomeres also prevent the ends of chromosomes from being recognised as sites of DNA damage. When telomeres become critically short, the DNA damage response may be triggered which can result in the inappropriate repair of telomeric DNA (Counter et al. 1992; Fagagna et al. 2003). This stage is referred to as telomere-driven crisis (Capper et al. 2007). DNA repair activities during crisis initiate events that contribute to genomic instability, including telomere fusions that may lead to the formation of dicentric chromosomes and the initiation of anaphase bridge breakage and fusion cycles (Gisselsson et al. 2000). This genomic instability can precipitate recombinations that result in the reactivation of telomerase, allowing the cell to escape this telomere-driven crisis (Halvorsen et al. 1999). This can result in the survival and proliferation of a clonal cancer cell with a highly rearranged genome. The ability of cells to escape crisis may be influenced by the type of fusion events which occur during crisis. Mammalian cells utilise different mechanisms to repair DSBs, including HR, C-NHEJ and A-NHEJ (Chang et al. 2017). The A-NHEJ pathway can result in the formation of intra-chromosomal fusion events such as sister chromatid fusions which are thought to be responsible for the deletion or amplification of specific chromosome arms, observed in most cancer types (Tabach et al. 2011; Cai and Sablina 2016). C-NHEJ may facilitate inter-chromosomal fusion events which although frequent, are thought to be insufficient to support escape from a telomere-driven crisis (Jones et al. 2014).

It has been demonstrated that the polymerase domain of POLQ is a crucial component of the A-NHEJ repair pathway, in which it promotes synapsis and inserts nucleotides at DSBs (Wyatt et al. 2016). Aside from its polymerase activity, POLQ has additional functions in DSB repair owing to the activity of its helicase domain (Ceccaldi et al. 2015). Previous data suggests that the helicase domain facilitates the interaction of POLQ with components involved in HR. For example, RPA binds to single-strand DNA (ssDNA) during the early stages of HR to stabilise the ssDNA (Pokhrel et al. 2017). The POLQ helicase domain has been shown to facilitate the removal of RPA from ssDNA, thus disrupting HR and promoting A-NHEJ, for which resected DNA is a substrate (as opposed to C-NHEJ which requires less resection to reveal microhomologies). In addition to this, it has been shown that POLQ can interact with RAD51 directly, displacing it from resected DSBs (Ceccaldi et al. 2015). The displacement of RAD51 also disrupts HR and the resected DNA is once again a substrate for A-NHEJ. The interaction of POLQ with both RPA and RAD51 therefore results in the inhibition of HR and increased levels of A-NHEJ (Mateos-Gomez et al. 2017).

### 3.3 AIMS OF THE CHAPTER

The purpose of this chapter was to ascertain whether the disruption of *POLQ* would impact on the ability of cells to escape telomere-driven crisis. This was achieved by driving clonal populations of HAP1 WT, HAP1 *POLQ*<sup>-</sup> and HCT116 WT, HCT116 *POLQ*<sup>-</sup> cells into a telomere-driven crisis. It was considered that the disruption of *POLQ* would reduce the frequency of inter-chromosomal end-end fusions and modulate the ability of cells to escape from a telomere-driven crisis.

The aims of the chapter were as follows:

- To induce telomere erosion and telomere-driven crisis in clonal populations of HAP1 and HCT116 WT and *POLQ*-deficient cells
- To document growth rates of HAP1 and HCT116 WT and *POLQ*-deficient cells as they progress through telomere-driven crisis
- To monitor the ability of WT and *POLQ*-deficient cells to escape crisis
- To determine whether the disruption of *POLQ* modulates telomere dynamics such as the rate of telomere erosion or telomere length heterogeneity

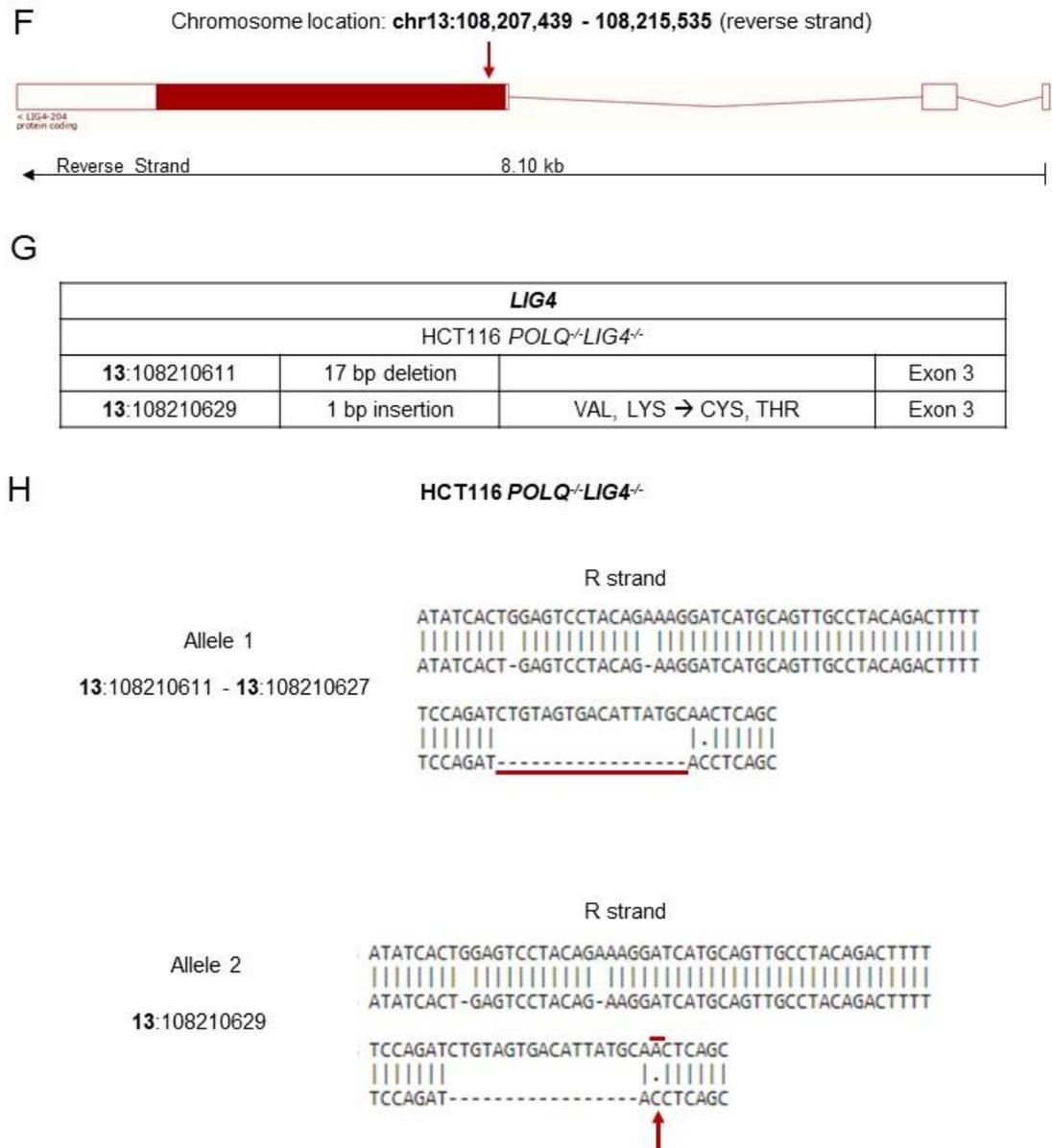
### 3.4 RESULTS

#### 3.4.1 Sequence verification of *POLQ* mutations

To assess the impact of DNA polymerase theta (*POLQ*) on DNA repair at telomeres and the ability of cells to escape a telomere-driven crisis, three human cell lines were used (HAP1, RPE1<sup>hTERT</sup> and HCT116) in which the *POLQ* gene had been disrupted using CRISPR/Cas9 editing. HAP1 human near-haploid cells (derived from KBM-7 leukaemic cells) were used in addition to HAP1 *POLQ*<sup>-</sup> cells that were obtained from Professor Geoffrey Higgins (Oxford University). Immortalised human primary retinal epithelial cells, RPE1<sup>hTERT</sup>, were used in parallel to RPE1<sup>hTERT</sup> *POLQ*<sup>-</sup> cells gifted from Professor Agnel Sfeir (Skirball Institute, New York University). HCT116 colorectal carcinoma cells were used to compare with HCT116 *POLQ*<sup>-</sup> and HCT116 *POLQ*<sup>-</sup>:*LIG4*<sup>-</sup> cells which were received from Professor Eric Hendrickson (University of Minnesota). Mutation at the CRISPR/Cas9 cut site was verified by DNA sequence analysis in all three cell lines. A 2 bp deletion at the CRISPR/Cas9-targeting sequence in exon 1 was observed in the HAP1 *POLQ*<sup>-</sup> cells. The same 1 bp deletion in exon 2 at the CRISPR/Cas9-targeting sequence was observed in both HCT116 and RPE1<sup>hTERT</sup> (**Figure 3.1A-D**).

HAP1 cells were used to reveal any differences in outcome following transition through a telomere-driven crisis in terms of cellular survival and telomere fusions induced by DNA repair mechanisms compared with a cancer derived diploid cell line (HCT116). RPE1<sup>hTERT</sup> cells were used to make comparisons between transformed cells with known and unknown pro-tumorigenic mutations (HAP1, HCT116) and non-cancerous cells. HCT116 *POLQ*<sup>-</sup>:*LIG4*<sup>-</sup> cells were used to analyse any difference in the abundance of fusion events induced by subtelomeric DSBs (*Chapter 4*) as both A-NHEJ and C-NHEJ are inhibited by the disruption of both *POLQ* and *LIG4* (**Figure 3.1E-H**).

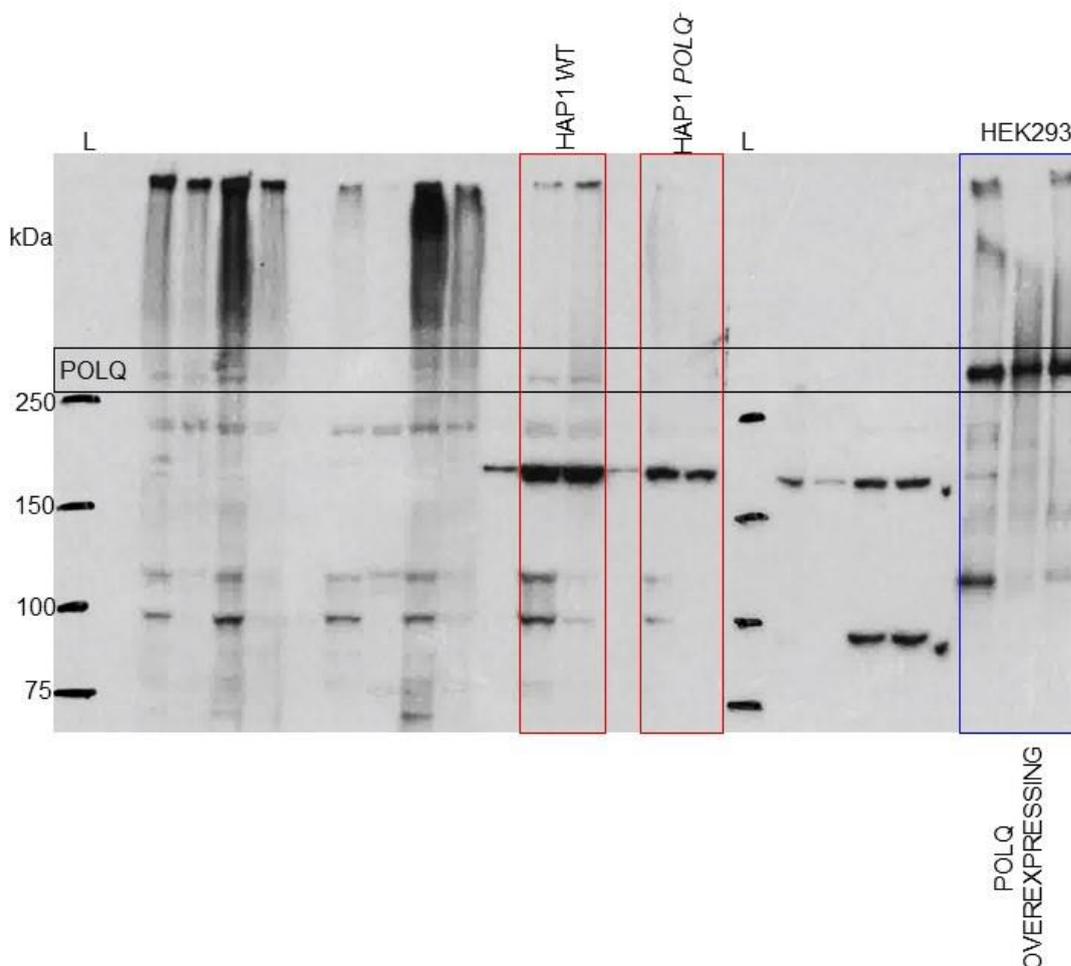




**Figure 3.1:** (A) Trace showing gene arrangement for human *POLQ* gene on the reverse stand of chromosome 3 (obtained from the Ensembl genome browser; GRCh38) Boxes represent exons. Lines connecting the boxes represent introns. Filled boxes represent coding sequence, and empty, unfilled boxes represent UTR (UnTranslated Region). Orange arrow indicates HAP1 CRISPR cut site, green arrow indicates HCT116 /RPE1 CRISPR cut site. (B) Table listing CRISPR targeting *POLQ* details for HAP1, HCT116 and RPE1 cell lines. (C) Sequence traces for HAP1 WT and HAP1 *POLQ*<sup>-/-</sup> cell lines with 2 bp deletion highlighted by boxes (forward strand (F strand) shown). (D) Sequence (top) indicating 1 bp deletion with green arrow for both HCT116 and RPE1 *POLQ*<sup>-/-</sup> cell lines (F strand shown). (E) Sequences indicating 1 bp insertion (top, R strand) and 1 bp deletion (bottom, F strand) targeting *POLQ* for HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line. (F) Trace showing gene arrangement for human *LIG4* gene on the R strand of chromosome 13 (obtained from the Ensembl genome browser; GRCh38). Red arrow indicates HCT116 CRISPR cut site. (G) Table listing CRISPR targeting *LIG4* details for HCT116 cell line. (H) Sequences indicating 17 bp deletion (top, R strand) and 1 bp insertion (bottom) targeting *LIG4* for HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line

### 3.4.2 Protein verification of *POLQ* mutations

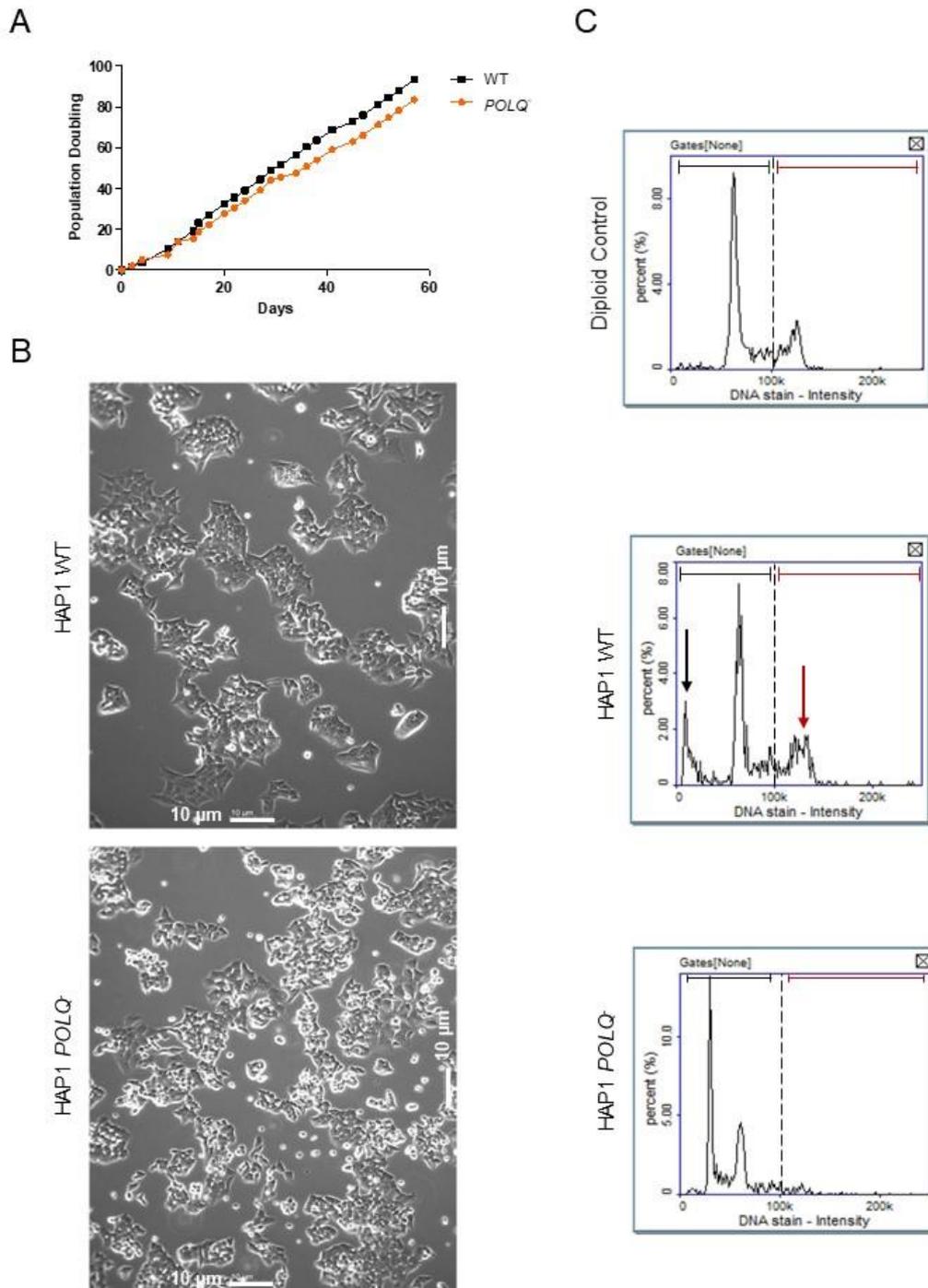
To confirm the disruption of *POLQ* resulted in decreased expression of *POLQ*, western blotting (*Chapter 2*) was performed using a monoclonal and a polyclonal anti-*POLQ* antibody. These antibodies were not specific enough to effectively confirm reduced expression of *POLQ*. As such, a commercially available anti-*POLQ* antibody could not be sourced. However, Professor Richard Wood (University of Texas, MD Anderson Cancer Center) has produced a mouse monoclonal anti-*POLQ* antibody that is currently still under development. Preliminary tests with the HAP1 cell line potentially demonstrate a reduction in the expression of *POLQ* in the *POLQ*-deficient cell line compared with the WT (**Figure 3.2**).



**Figure 3.2:** Western blot for HAP1 WT and HAP1 *POLQ*<sup>-</sup> cell lines (red box) alongside Human Embryonic Kidney (HEK) 293 cells overexpressing *POLQ* (blue box) using a mouse monoclonal anti-*POLQ* antibody (Atb5). Experiment performed by Megan Lowery. Data received from Professor Richard Wood, University of Texas.

### 3.4.3 HAP1 Cell Line

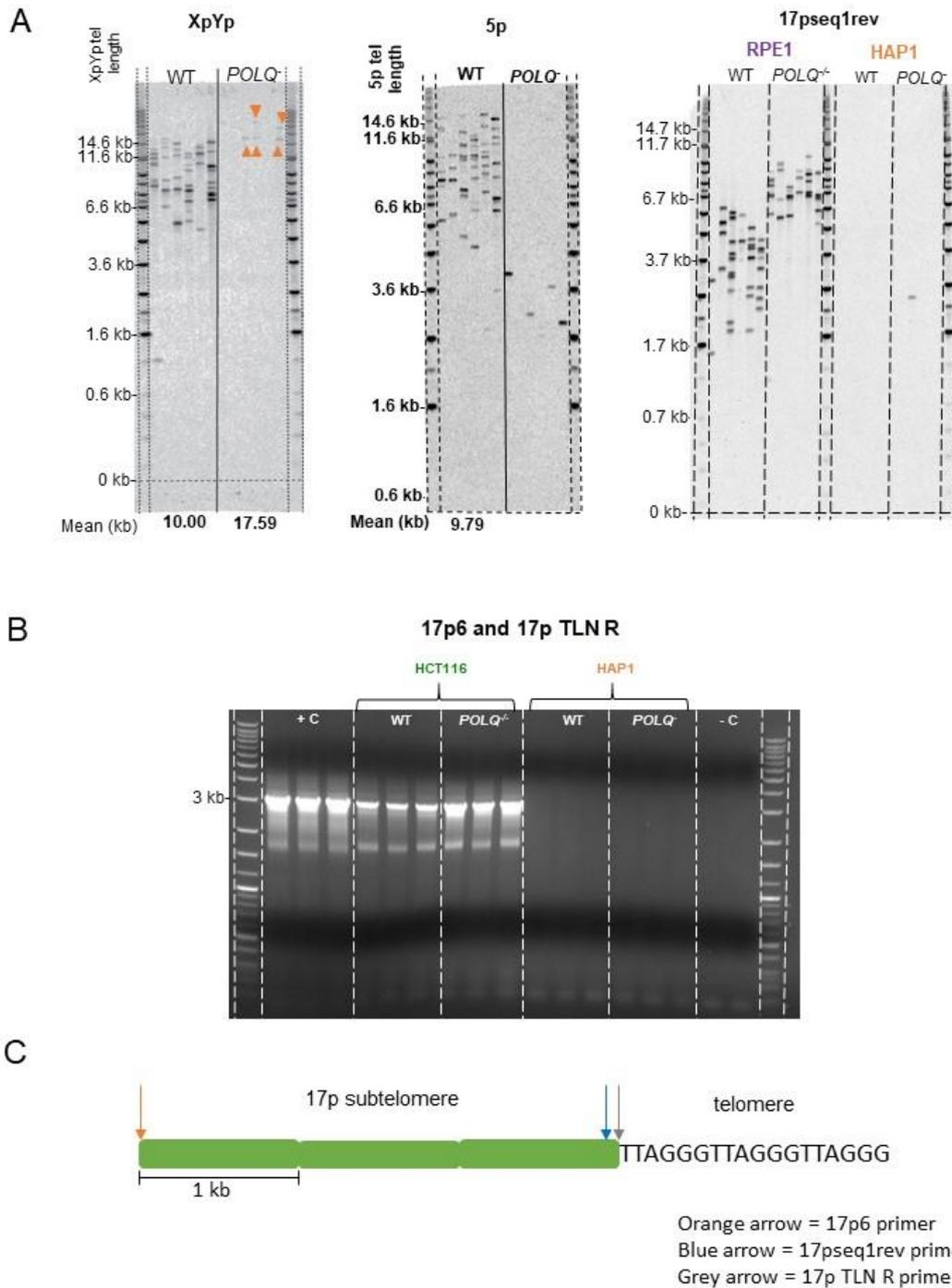
The disruption of *POLQ* did not have a significant impact on the growth rate of HAP1 cells when compared with HAP1 WT cells. HAP1 WT and *POLQ*<sup>-</sup> cell lines were cultured for 57 days and completed 93 and 84 population doublings (PDs) respectively. The mean doubling rate for the HAP1 WT cell line was 1.72 PDs per day compared to a mean doubling rate of 1.61 PDs per day for the HAP1 *POLQ*<sup>-</sup> cell line ( $p=0.3560$ , Mann Whitney test) (**Figure 3.3A**). There was no observable difference in the morphology of the HAP1 *POLQ*<sup>-</sup> cell line compared to the HAP1 WT cell line using microscopy at 4x (not shown) or 10x magnification (**Figure 3.3B**). Cell samples were fixed with 70% ethanol in order to stain DNA with 4',6-diamidino-2-phenylindole (DAPI). Image cytometry was used to visualise the relative DNA content of the cell lines and thus determine the ploidy of the parental lines. DAPI binds DNA stoichiometrically so the stain intensity is directly proportional to DNA content. The HAP1 WT parental cell line produced peaks that represent diploid cells (1N and 2N DNA content, calibrated by diploid comparators), whilst the HAP1 *POLQ*<sup>-</sup> parental cell line produced peaks that indicated the cells were haploid. However, an additional peak was observed for the HAP1 WT parental cell line that does not align with either 1N or 2N DNA content. This additional peak may reflect the culture undergoing diploidisation, during which some cells have ploidy changes at some chromosomes. Alternatively, this additional peak could represent mitochondrial DNA (**Figure 3.3C**).



**Figure 3.3:** (A) Growth curves (PD per day) for HAP1 WT and HAP1 *POLQ*<sup>-</sup> over 57 days in culture. (B) Microscope photographs of HAP1 WT (top) and HAP1 *POLQ*<sup>-</sup> cells (bottom) at 10x magnification. Scale bar highlighted in white. (C) Plots demonstrating DAPI staining of DNA to show ploidy of the HAP1 WT (middle) and HAP1 *POLQ*<sup>-</sup> parental cell lines (bottom) compared to a diploid control (HCT116) (top). One example shown, analysis repeated in triplicate. Peaks present at a stain intensity of 100 k or below represents haploid DNA content (2N) whereas peaks with a stain intensity above 100 k represent diploid DNA content (4N). Diploid population obtained with HAP1 WT cells indicated by a red arrow, and an additional peak indicated by a black arrow.

### Telomere Lengths in HAP1 cell line

The XpYp telomere length of the HAP1 parental lines was analysed by single telomere length analysis (STELA - a single-molecule long-range PCR) and showed that the parental HAP1 *POLQ*<sup>-</sup> cell line had a mean XpYp telomere length of 17.59 kb compared to a significantly shorter mean of 10.00 kb for the WT parental line ( $p < 0.001$ , Mann Whitney test). Measuring the telomere lengths of additional chromosome ends allowed comparisons to be drawn between different telomeres in each cell line. At 9.79 kb, the mean length calculated for the 5p telomere of the HAP1 WT cells was very similar to their mean XpYp telomere length. However, there were few bands present on the 5p STELA blot for the HAP1 *POLQ*<sup>-</sup> line. Therefore, the mean length of the 5p telomere could not be accurately calculated using STELA for this cell line. This may be because the 5p specific subtelomere primer site does not exist in this cell line or because the length of the 5p telomeres were beyond the detectable range for STELA. The long telomere lengths of the HAP1 *POLQ*<sup>-</sup> cell line was further demonstrated by a terminal restriction fragment (TRF) assay which showed that this cell line had a mean telomere length of 21.9 kb. This is a mean of all chromosome ends, rather than a mean of repeated measure at one telomere (*Section 3.4.6*). Standard 17p STELA conditions (17pseq1rev primer, 0.31 kb from the telomere) failed to amplify telomeric molecules in the HAP1 cell line (**Figure 3.4A**). I therefore sequenced the 17p subtelomere in this cell line in order to design a bespoke assay for measuring 17p telomere length in HAP1 cells. I used the 17p6 (forward) primer 3 kb from the start of the 17p telomere with the 17p TALEN R (reverse) primer, which is close to the telomere, to amplify 17p subtelomere, producing a 3074 bp amplicon. However, this PCR was also unsuccessful, despite working in HCT116 and Human Embryo Kidney (HEK293) cell lines (**Figure 3.4B**). These data suggest that the HAP1 cell lines have a different 17p subtelomere sequence and structure (**Figure 3.4C**).



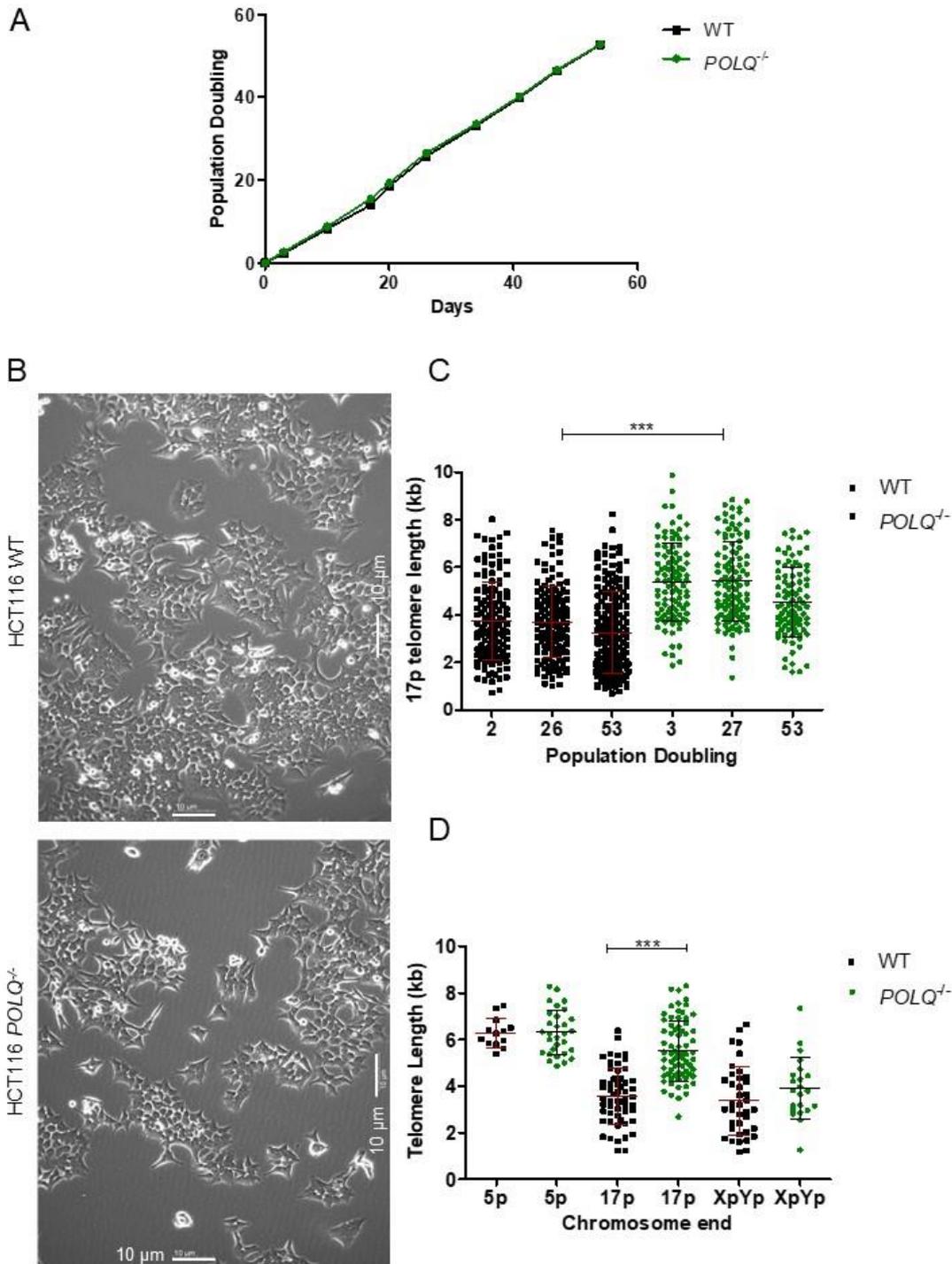
**Figure 3.4:** (A) STELA profile of HAP1 parental lines using the XpYp (left) and 5p (middle) primer. STELA profile for RPE1<sup>hTERT</sup> (positive control) and HAP1 parental lines using the 17pseq1rev primer (left). 6 reactions were repeated for each sample. (B) Gel image of 3074 bp amplicon generated using 17p6 and 17p TLN R primers to amplify the 17p subtelomere in HCT116 and HAP1 cell lines, including positive (HEK293 cells) and negative controls. (C) Diagram representing 17p subtelomere with 17p6 primer location highlighted by orange arrow, 17pseq1rev primer location highlighted by blue and 17p TALEN reverse (TLN R) primer location highlighted by grey arrow.

### 3.4.4 HCT116 Cell Line

As observed with the HAP1 cell line, the disruption of *POLQ* in HCT116 cells did not inhibit cell growth. HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cell lines were cultured for 54 days and completed 52.7 and 52.8 PDs, respectively. The mean doubling rate for both the HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cell lines was 1.00 PD per day ( $p=0.7572$ , Mann Whitney test). (**Figure 3.5A**). Similarly, there was no observable difference in the morphology of the HCT116 *POLQ*<sup>-/-</sup> cell line compared with the HCT116 WT cell line at 4x (not shown) or 10x magnification (**Figure 3.5B**). STELA was used to determine whether there were any differences in telomere dynamics for these cell lines following >50 PDs in culture. STELA at an early (PD 2-3), mid (PD 26-27) and a late (PD 53) time point showed no significant change in 17p telomere length distribution in both the HCT116 WT ( $p=0.9608$ , Mann Whitney test) and HCT116 *POLQ*<sup>-/-</sup> ( $p=0.9602$ , Mann Whitney test) cell lines due to the maintenance of telomere length via the activity of telomerase (*Figure 3.14F*, *Section 3.4.8*). The HCT116 WT line maintained a mean 17p telomere length of 3.56 kb compared to a significantly longer mean 17p telomere length of 5.10 kb for the HCT116 *POLQ*<sup>-/-</sup> cell line ( $p<0.001$ , Mann Whitney Test) (**Figure 3.5C**).

#### Telomere Lengths in HCT116 cell line

Although the mean 17p telomere length calculated for the HCT116 WT was significantly shorter than the mean 17p telomere length of the HCT116 *POLQ*<sup>-/-</sup> line ( $p<0.001$ , Mann Whitney test), the differences in lengths of the XpYp and the 5p telomere for the HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cell lines were not statistically significant. The mean XpYp telomere length calculated for the HCT116 WT cell line was 3.30 kb compared to 3.76 kb for the HCT116 *POLQ*<sup>-/-</sup> cell line. Similarly, the mean 5p telomere length calculated for the HCT116 WT cell line was 6.29 kb compared to 6.31 kb calculated for the HCT116 *POLQ*<sup>-/-</sup> cell line (**Figure 3.5D**).



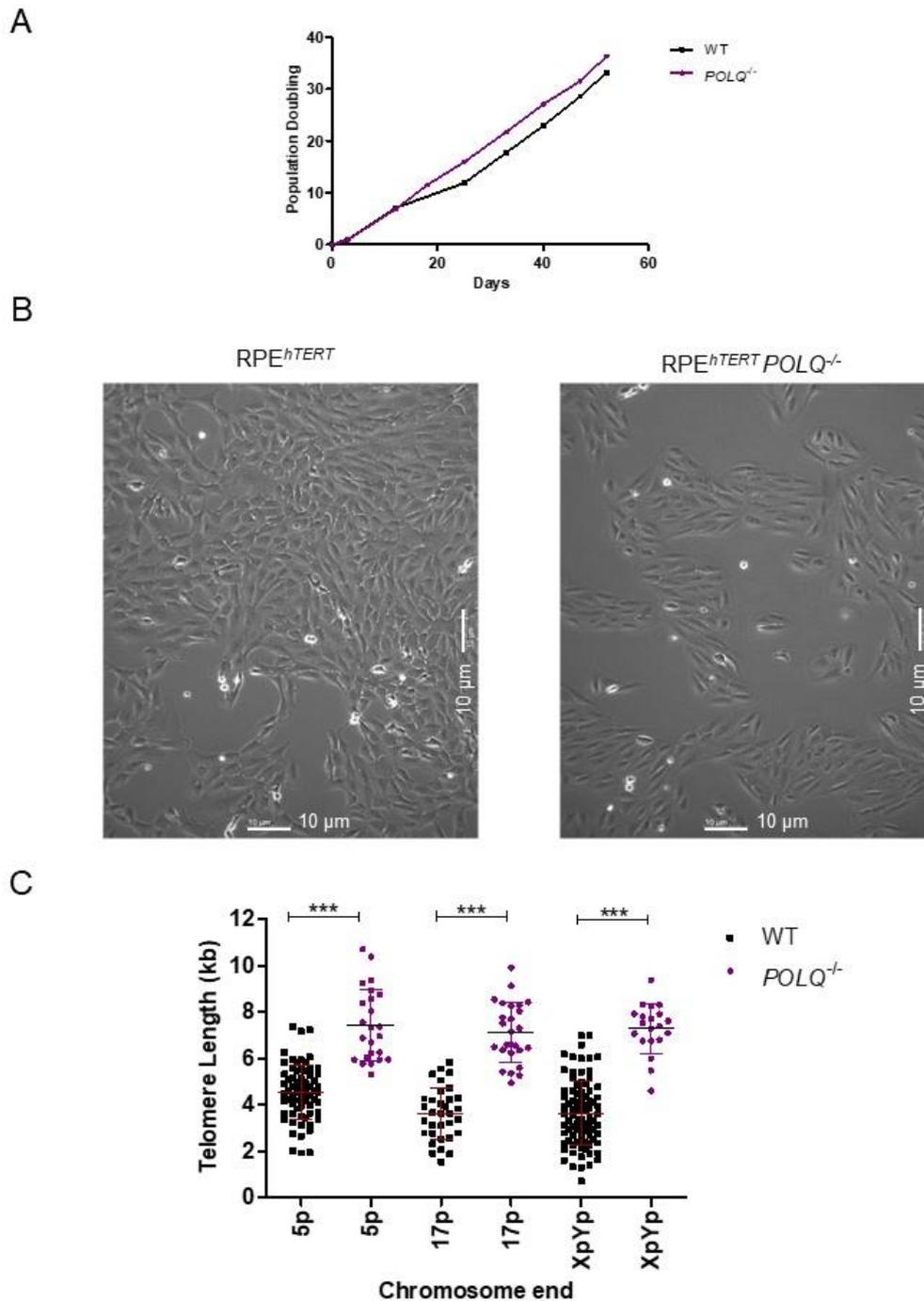
**Figure 3.5:** (A) Growth curves for HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cell lines over 54 days. (B) Microscope photograph of HCT116 WT (top) and HCT116 *POLQ*<sup>-/-</sup> cells (bottom) at 10x magnification. Scale bar highlighted in white. (C) Graphical representation of 17p STELA profiles produced for HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> parental cell lines for three different population doubling time points. Mean telomere length is shown in red (HCT116 WT) and black (HCT116 *POLQ*<sup>-/-</sup>) alongside standard deviation intervals. Significance was determined using a Mann Whitney test. 6 reactions were performed for each sample. (D) Graphical representation of STELA profiles for HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cell lines (PD 2 and 3 respectively) at 5p, 17p and XpYp chromosome ends. 6 reactions were performed for each sample. Mean telomere length is shown in red (HCT116 WT) and black (HCT116 *POLQ*<sup>-/-</sup>) alongside standard deviation intervals. Significance was calculated using Mann Whitney test.

### 3.4.5 RPE1 Cell Line

The disruption of *POLQ* in RPE1<sup>hTERT</sup> cells did not affect cell growth. RPE1<sup>hTERT</sup> WT and RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup> cell lines were cultured for 52 days and completed 33.2 and 36.3 PDs, respectively. The mean doubling rate for the RPE1<sup>hTERT</sup> WT cell line was 0.71 PDs per day compared to 0.73 PDs per day for the RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup> cell line ( $p=0.9078$ , Mann Whitney test) (**Figure 3.6A**). As observed with the HAP1 and HCT116 cell lines, there was also no detectable difference in the morphology of the RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup> cell line compared to the RPE1<sup>hTERT</sup> WT cell line at 4x (not shown) or 10x magnification (**Figure 3.6B**).

### Telomere Lengths in the RPE1 cell line

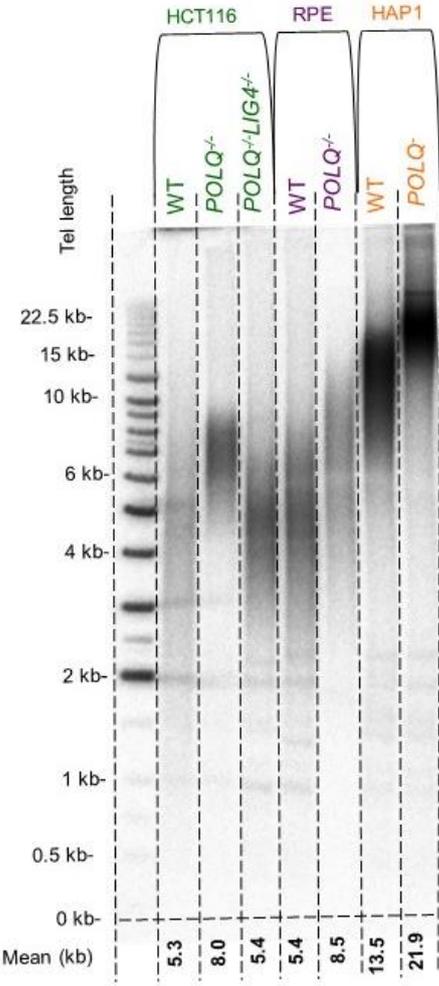
The disparity in telomere lengths between the WT and *POLQ*-deficient cell lines was also observed in the RPE1<sup>hTERT</sup> cell line, as the RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup> parental cell line displayed longer telomeres at three different chromosome ends analysed compared with the RPE1<sup>hTERT</sup> WT cell line. For the RPE1<sup>hTERT</sup> WT line, the mean XpYp telomere length was 3.59 kb, the mean 17p telomere length was 3.62 kb and the mean 5p telomere length was slightly longer at 4.54 kb. For the RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup> line, the mean XpYp telomere length was 6.77 kb, the mean 17p telomere length was 7.09 kb and the mean 5p telomere length was 7.42 kb. When comparing all three chromosome ends, the RPE1<sup>hTERT</sup> WT cell line has significantly shorter telomere lengths than the RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup> cell line ( $p<0.001$ , Mann Whitney test) (**Figure 3.6C**).



**Figure 3.6:** **(A)** Growth curves for RPE1  $WT^{hTERT}$  and RPE1  $POLQ^{-/-}:hTERT$  cell lines over 52 days. **(B)** Microscope photograph of RPE1  $WT^{hTERT}$  (left) and RPE1  $POLQ^{-/-}:hTERT$  cells (right) at 10x magnification. Scale bar highlighted in white. **(C)** Graphical representation of STELA profiles produced for RPE1  $WT^{hTERT}$  and RPE1  $POLQ^{-/-}:hTERT$  parental cell lines at 5p, 17p and XpYp chromosomes. 6 reactions were performed for each sample. Mean telomere length is shown in red (RPE1  $WT^{hTERT}$ ) or black (RPE1  $POLQ^{-/-}:hTERT$ ) alongside standard deviation intervals. Significance was determined using a Mann Whitney test.

### 3.4.6 Comparison of telomere lengths for WT and POLQ-deficient cell lines

Despite differences in mean telomere lengths between the parental cell lines (HAP1, HCT116 and RPE1<sup>hTERT</sup>), increased telomere lengths were recorded for all three parental POLQ-deficient cell lines compared with the parental WT cell lines. A TRF assay was performed to determine whether this length differential was observed at all chromosome ends. The TRF assay demonstrated that this telomere length differential between WT and POLQ-deficient parental lines was observed across all chromosomes. The HAP1 WT parental cell line had a mean telomere length of 13.5 kb compared to a mean telomere length of 21.9 kb for the HAP1 *POLQ*<sup>-/-</sup> parental line. Likewise, HCT116 WT cells had a shorter mean telomere length of 5.3 kb compared to a mean telomere length 8.0 kb for the HCT116 *POLQ*<sup>-/-</sup> parental line. In addition, the RPE1<sup>hTERT</sup> *POLQ*<sup>-/-</sup> parental cell line had a longer mean telomere length of 5.4 kb compared to a mean telomere length 8.5 kb for the RPE1 WT<sup>hTERT</sup> cell line (**Figure 3.7**).



**Figure 3.7:**TRF blot for parental cell lines (PD 2/3). Mean telomere lengths listed under each sample. TRF analysis completed once due to large quantities of DNA required.

**Table 3.1: Summary of telomere length profiles obtained for parental cell lines (PD 2/3) used in this study**

Cell Line	Mean telomere length (kb)			
	STELA			TRF
	5p	17p	XpYp	All chromosome ends
HAP1 WT	9.79	N/A	10.00	13.5
HAP1 <i>POLQ</i> <sup>-</sup>	N/A	N/A	17.59	21.9
HCT116 WT	6.29	3.56	3.30	5.3
HCT116 <i>POLQ</i> <sup>-</sup>	6.31	5.10	3.76	8.0
RPE1 <sup>hTERT</sup> WT	4.54	3.62	3.59	5.4
RPE1 <sup>hTERT</sup> <i>POLQ</i> <sup>-</sup>	7.42	7.09	6.77	8.5

### 3.4.7 A Telomere-Driven Crisis in the HAP1 Cell Line

Telomere lengths of immortalised HAP1 WT and HAP1 *POLQ*<sup>-</sup> cells are maintained by telomerase activity (*Figure 3.12, Section 3.4.7.3*), hence abrogation of telomerase activity is required to induce telomere erosion and initiate a telomere-driven crisis. To disrupt telomerase activity, cells were transduced with amphotropic retroviral vectors (pBABE-puro) containing a dominant-negative telomerase reverse transcriptase construct (DN-hTERT) and puromycin resistance-encoding sequence. DN-hTERT inhibits telomerase via two different mechanisms: by facilitating the degradation of wild type hTERT through dimerization and by competitive inhibition of the interaction between the wild type hTERT and the TERC subunit (Nguyen et al. 2009). Control cells (HAP1 WT<sup>PURO</sup> and HAP1 *POLQ*<sup>-PUR0</sup>) were generated that express only the puromycin-resistance cassette driven by the pBABE-puro vector. This allowed the assessment of the background impact of retroviral transduction and acquisition of puromycin resistance for telomere length stability in comparison with the deliberate targeting of telomerase (hTERT) activity. Puromycin selection was initiated 24 hr post-infection to select for cells that had successfully incorporated the DN-hTERT or the empty pBABE-puro vector. Cloning rings were used to pick single cell clones (10

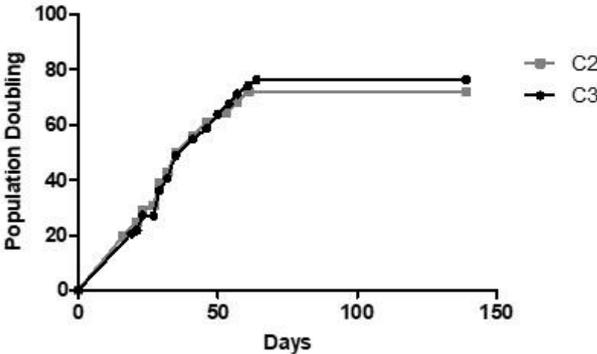
HAP1 WT<sup>DN-hTERT</sup>, 10 HAP1 *POLQ*<sup>DN-hTERT</sup>, 5 HAP1 WT<sup>PURO</sup> and 5 *POLQ*<sup>-:Puro</sup> clones) and population growth was monitored for each clone after transduction. Following DN-hTERT transduction, cells were cultured and sampled at regular intervals throughout their replicative lifespan. Telomere lengths were analysed using STELA at multiple sampling points and compared with those obtained at the first passage post-transduction. Reductions in telomere length allowed verification of the functional incorporation of the DN-hTERT cassette that suppresses telomerase activity.

### 3.4.7.1 HAP1 Clone Survival: Senescent Clones

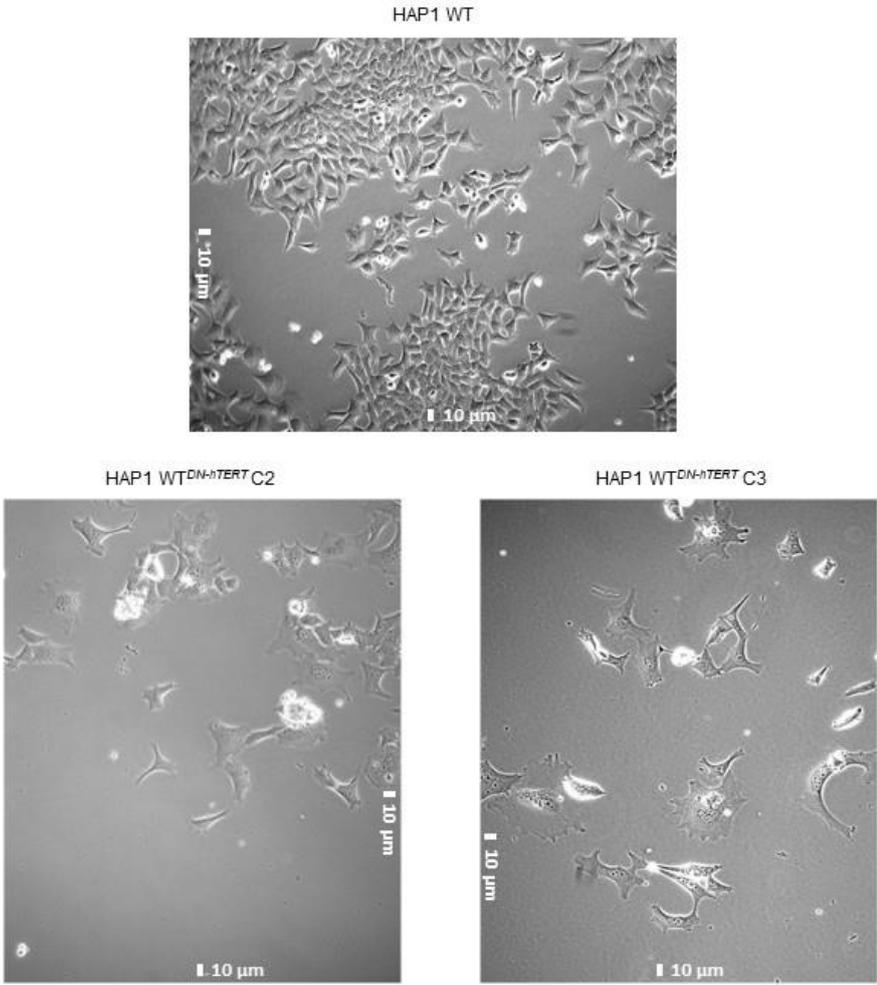
Single cell clones were continually cultured through the onset of, and the escape from, telomere-driven crisis. Of the 10 HAP1 WT<sup>DN-hTERT</sup> clones that were maintained in culture, there was an 80% survival rate, compared to a 100% survival rate for the HAP1 POLQ<sup>DN-hTERT</sup> clones. The two WT clones (C2 and C3) that did not survive grew in culture for 61 and 64 days, respectively, completing 72 and 76 PDs during this period (**Figure 3.8A**). Following this, these clones entered a period of replicative arrest and increased cell size (consistent with cells undergoing replicative senescence) (**Figure 3.8B**).

STELA analysis of the two clones that entered senescence showed that these clones had mean starting XpYp telomere lengths of 4.42 kb and 5.92 kb, respectively. These clones therefore had considerably shorter XpYp telomere lengths than the parental cell line (9.93 kb). Erosion of the XpYp telomere was observed as a function of increased PDs, until a mean telomere length of 1.22 kb and 1.86 kb was reached, following which cell division for these clones ceased (**Figure 3.8C**). This telomere erosion associated with the abrogation of telomerase was also observed at 9p, 16p and 11q chromosome ends. Erosion of the 11q telomere for C3 was observed until a mean 11q telomere length of 0.92 kb was reached. Likewise, erosion of the telomeres on chromosome 9p and 16p was seen for C3 until mean telomere lengths of 1.59 kb and 2.08 kb were reached (**Figure 3.8D**).

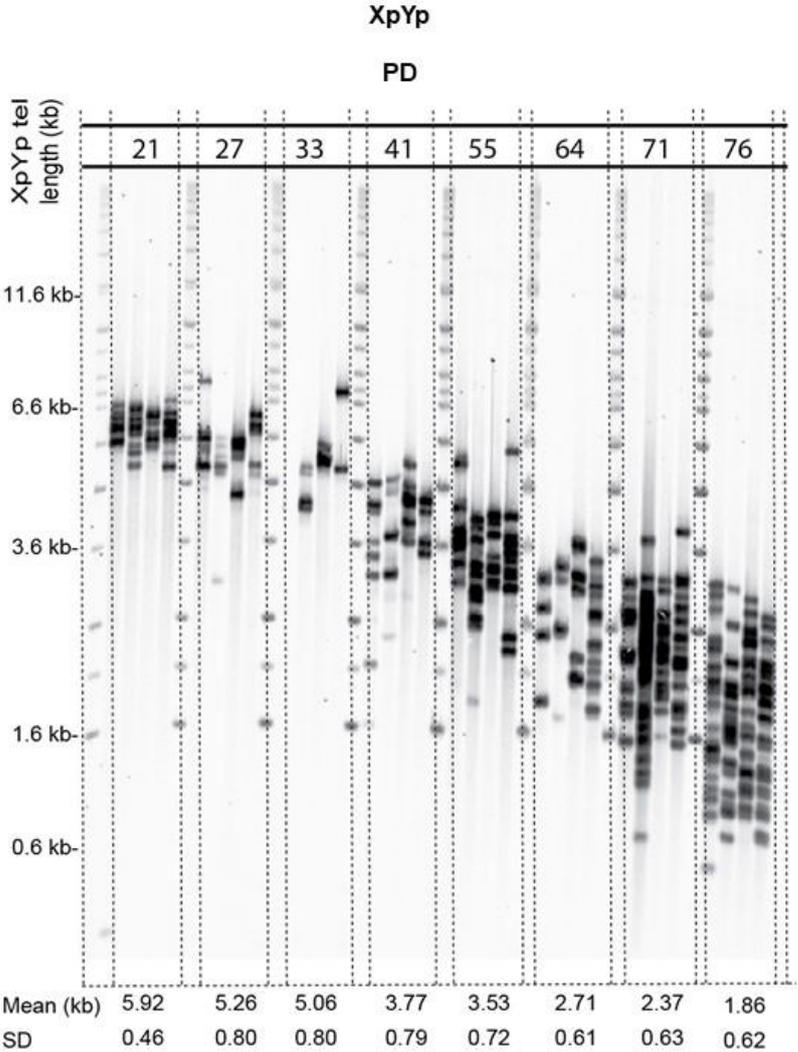
A



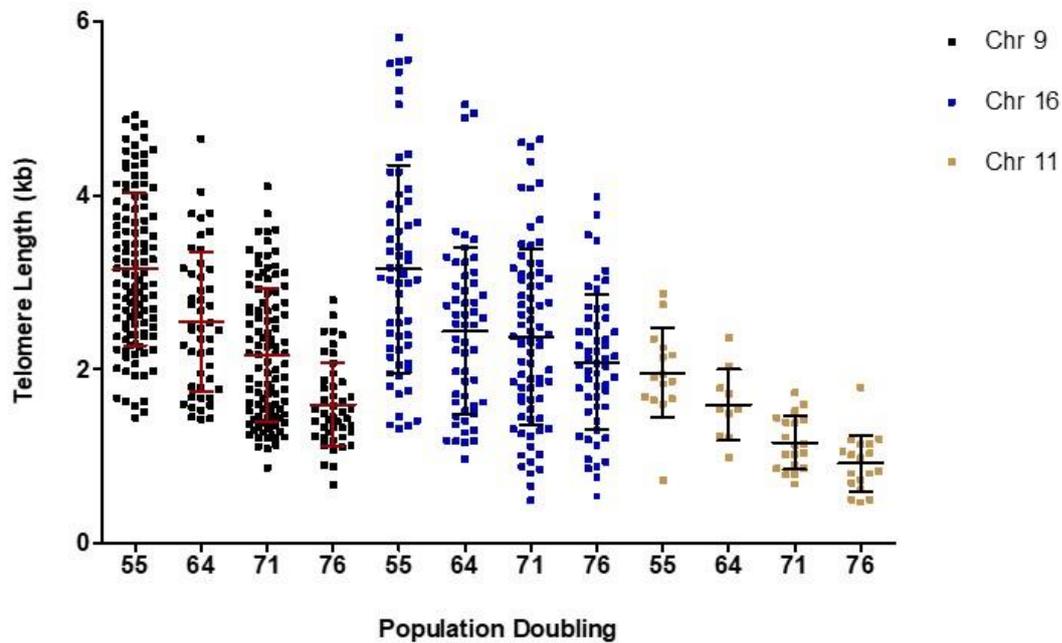
B



C



D



**Figure 3.8:** (A) Growth curves for HAP1 WT<sup>DN-hTERT</sup> C2 and C3 (PD/day). (B) Microscope photographs of C2 (right) and C3 (left) at 10x magnification. Scale bar highlighted in white. (C) STELA profile showing erosion of XpYp telomere for HAP1 WT<sup>DN-hTERT</sup> C3. 4 reactions performed for each sample. (D) Graphical representation of the STELA profiles for HAP1 WT<sup>DN-hTERT</sup> C3 at 9p, 16p and 11q telomeres. 4 reactions performed for each sample. The mean telomere length for each time point is shown in black (16p/11q) or red (9p) alongside standard deviation intervals

### 3.4.7.2 HAP1 Surviving Clones: Telomere-driven crisis

The eight surviving HAP1 WT<sup>DN-hTERT</sup> clones completed 1.26 PDs per day (mean) except for C13 which grew more slowly than all other HAP1 WT<sup>DN-hTERT</sup> clones, completing an average of 0.88 PDs per day. HAP1 WT<sup>DN-hTERT</sup> clones were cultured for 183 – 231 days and underwent between 162 and 242 PDs before the cultures were intentionally terminated. HAP1 POLQ<sup>-DN-hTERT</sup> clones completed 1.27 PDs per day (mean), except one clone (C8) which grew more slowly and underwent 0.85 PDs per day (mean). HAP1 POLQ<sup>-DN-hTERT</sup> clones continued to grow in culture for up to 248 days following transduction, completing between 180 and 278 PDs before the cultures were intentionally terminated. Slowed population growth associated with telomere-driven crisis was observed for the HAP1 WT<sup>DN-hTERT</sup> clones. This cell line had a period of crisis (defined by slowed growth rates) for 12 days (mean). In contrast, the HAP1 POLQ<sup>-DN-hTERT</sup> clones had a period of crisis for 9 days (mean). No change in growth kinetics were observed for either the HAP1 WT<sup>PURO</sup> or HAP1 POLQ<sup>-Puro</sup> clones (one of each clone cultured for the duration of the experiment). HAP1 WT<sup>PURO</sup> C1 completed on average 1.68 PDs per day and therefore completed 234 PDs in total over 144 days in culture. HAP1 POLQ<sup>-Puro</sup> C1 completed on average 1.60 PDs per day, completing 222 PDs in total over 144 days in culture before the cultures were terminated (**Figure 3.9A**).

#### Starting telomere lengths

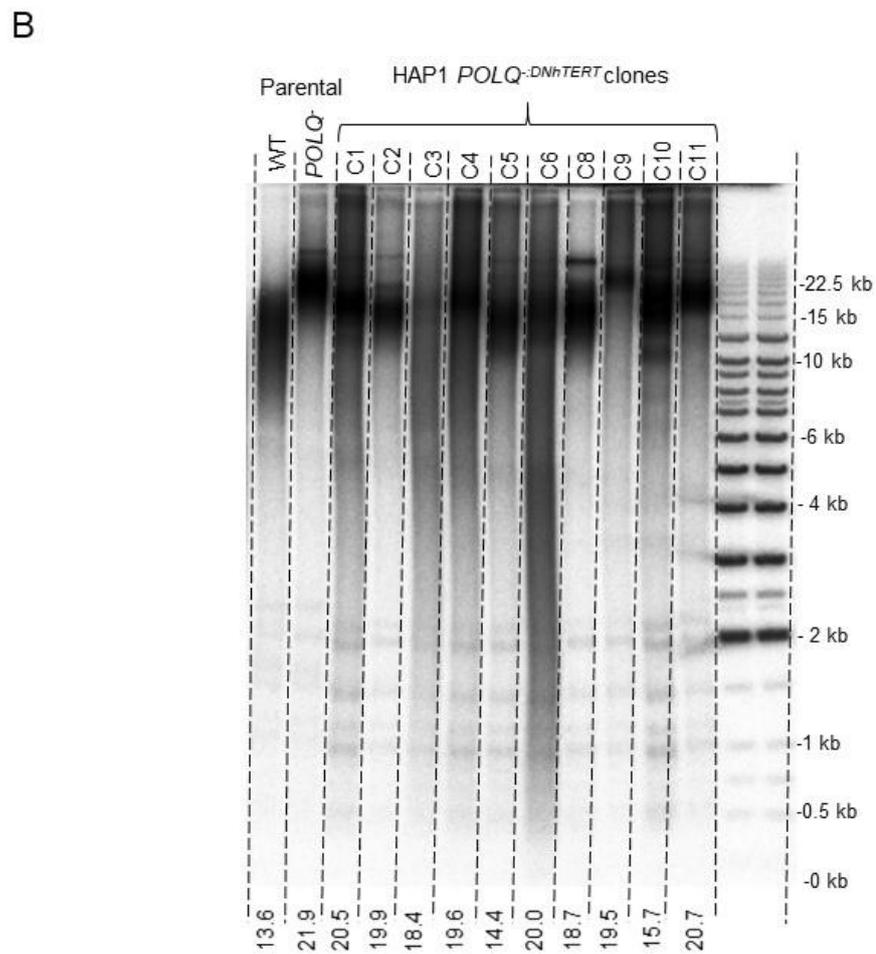
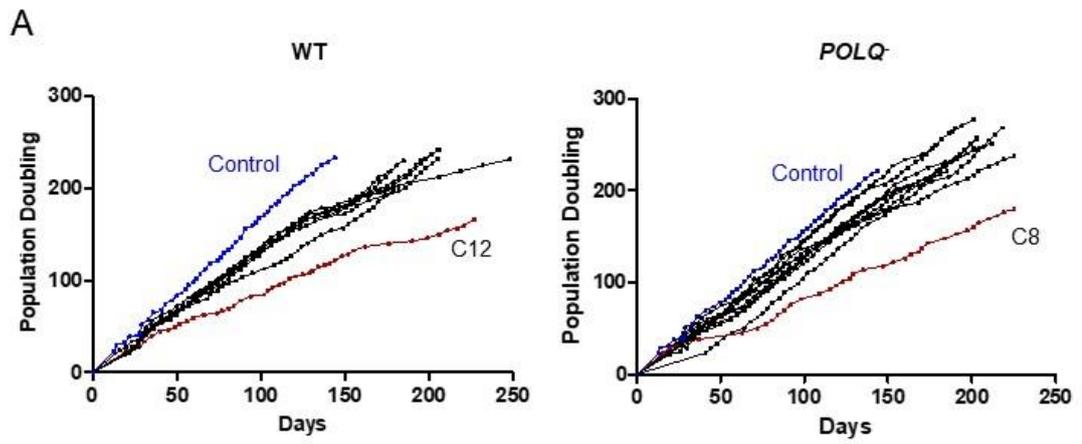
The starting XpYp telomere lengths of the HAP1 WT<sup>DN-hTERT</sup> clones following transduction (~PD 20) was between 5.54 kb and 8.22 kb (mean 6.52 kb). However, the telomere lengths of the HAP1 POLQ<sup>-DN-hTERT</sup> clones could not be effectively detected as their telomeres were beyond the detectable length range of STELA. Between 22 and 34 PDs were required before the HAP1 POLQ<sup>-DN-hTERT</sup> clones had undergone sufficient telomere erosion such that their telomere lengths were more tractable for amplification and detection using STELA. At this stage, these clones had

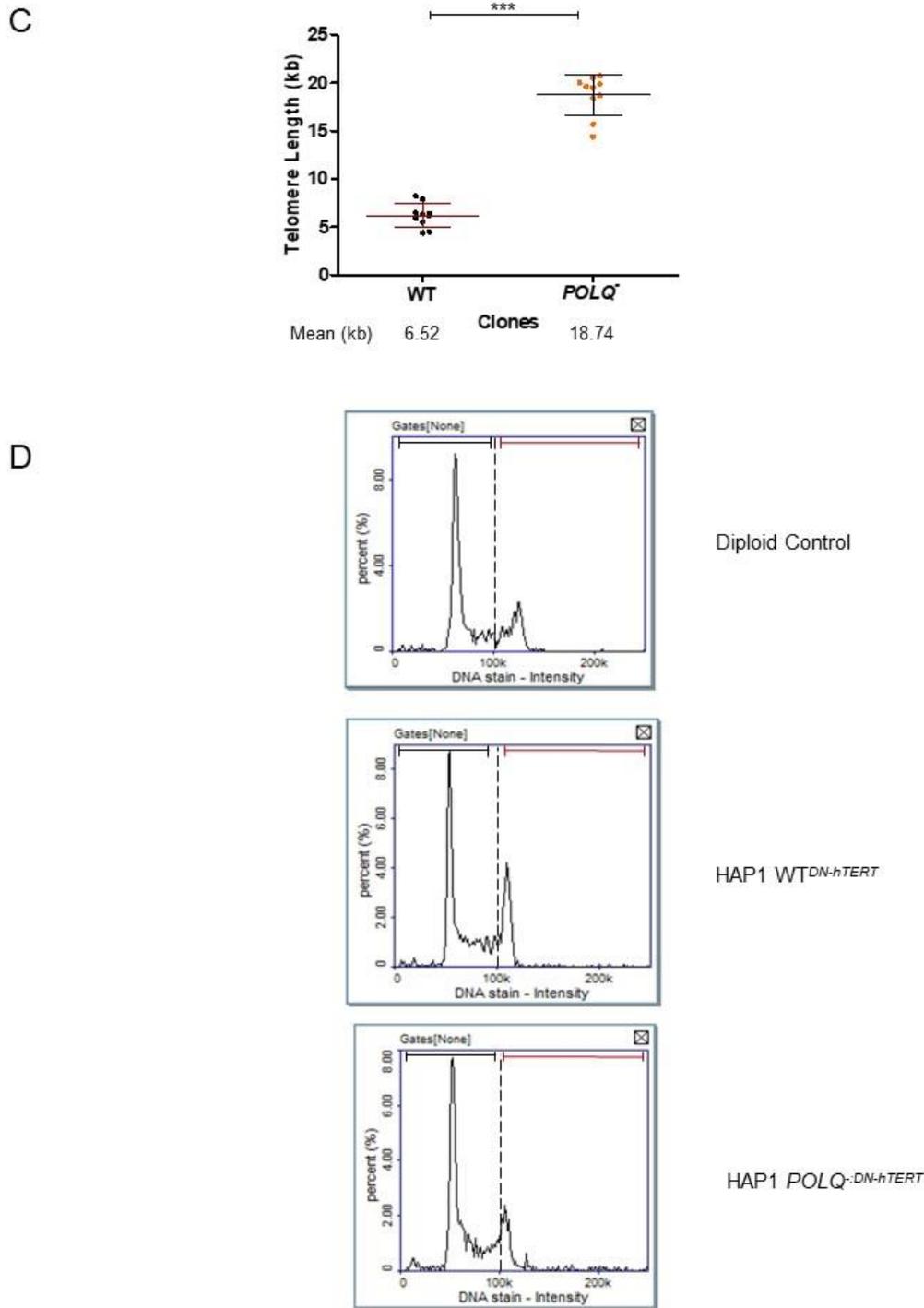
XpYp telomere lengths between 9.37 kb and 15.08 kb (mean 11.92 kb). For two clones, 92 and 103 PDs were required for telomere lengths to be detected using STELA, at which stage these clones had XpYp telomere lengths of 10.13 kb and 13.16 kb. A TRF assay was performed for the early HAP1 *POLQ<sup>-DN-hTERT</sup>* samples to determine overall mean telomere lengths for these clones. This assay revealed that these clones had mean starting telomere lengths of 18.74 kb, but this accounts for all chromosomes so is likely to be different from the unique telomere lengths measured by STELA (**Figure 3.9B-C**).

In contrast, STELA at a mid (PD 95) and late time point (PD 234) for the HAP1 *WT<sup>PURO</sup>* control demonstrated a mean (unchanged) XpYp telomere length of 5.87 kb and 5.72 kb respectively, consistent with the absence of the DN-hTERT cassette.

### **Changes in ploidy**

As these cells had been cultured for an extensive period, DNA staining with DAPI was repeated in HAP1 *DN-hTERT* clones to ascertain whether these clones had become diploid over time. Image cytometry analysis revealed that both HAP1 *WT<sup>DN-hTERT</sup>* and HAP1 *POLQ<sup>-DN-hTERT</sup>* clones had become diploid before transiting crisis (**Figure 3.9D**).





**Figure 3.9:** (A) Growth curves for HAP1 WT<sup>DN-hTERT</sup> (left), HAP1 *POLQ*<sup>-DN-hTERT</sup> and (right) HAP1<sup>PURO</sup> clones (highlighted in blue). (B) TRF blot for HAP1 parental cell lines (left) and HAP1 *POLQ*<sup>-DN-hTERT</sup> clones at early time points (PD2/3) (right). (C) Graphical representation of mean starting telomere lengths for HAP1 WT<sup>DN-hTERT</sup> clones at the XpYp telomere and starting telomere lengths for HAP1 *POLQ*<sup>-DN-hTERT</sup> obtained from TRF. Mean telomere length is shown in red (WT) and black (*POLQ*<sup>-</sup>) alongside standard deviation intervals. (D) Graph depicting DAPI DNA stain intensity for the diploid control (HCT116) (top), HAP1 WT<sup>DN-hTERT</sup> (middle) and HAP1 *POLQ*<sup>-DN-hTERT</sup> clones (bottom) shows diploid DNA content (4N DNA content) One example for each cell line shown.

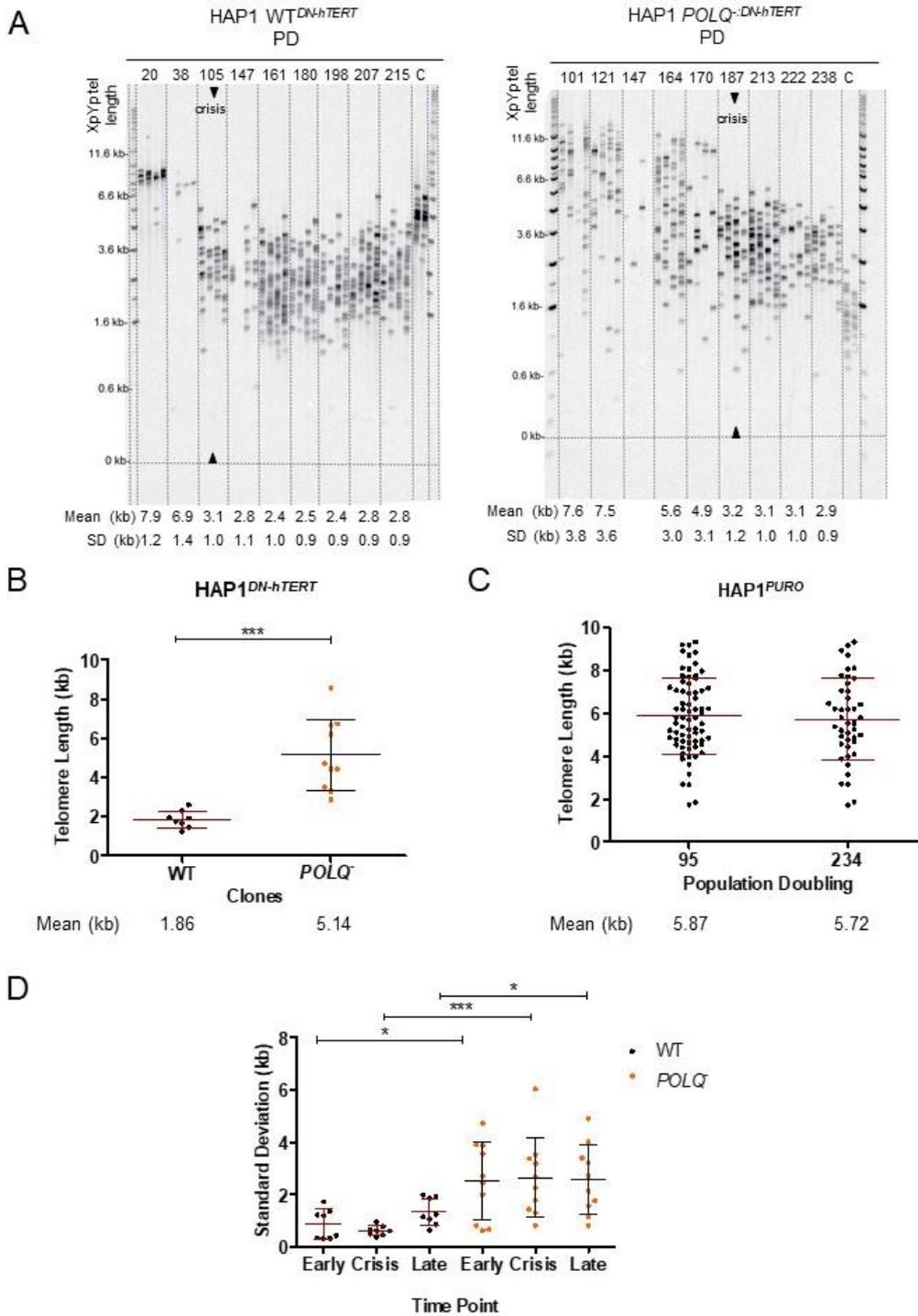
### 3.4.7.3 Telomere Length Analysis in HAP1<sup>DN-hTERT</sup> clones

STELA of the HAP1 WT<sup>DN-hTERT</sup> clones demonstrated telomere erosion to a mean XpYp telomere length of 1.86 kb (shortest telomere length representing initiation of telomere-driven crisis). Following this, telomere lengths were maintained at a mean length of 2.52 kb for 5 out of the 8 clones. However, XpYp telomere lengths for the remaining 3 out of the 8 clones continued to increase to lengths >3.00 kb.

STELA of the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones demonstrated XpYp telomere erosion for all clones, although the telomere length at which erosion ceased was not consistent and varied from 3.05 kb to 9.52 kb (mean 5.74 kb) (**Figure 3.10A**). There was, therefore, variation in the shortest XpYp telomere lengths detected for these clones, ranging from 2.87 kb to 8.55 kb (mean 5.14 kb) (**Figure 3.10B**).

In contrast, no significant change was observed for XpYp telomere lengths over time for the HAP1 WT<sup>PURO</sup> clone ( $p=0.7611$ , Mann Whitney test) (**Figure 3.10C**).

An increase in the heterogeneity of XpYp telomere length profiles following crisis was observed in the WT<sup>DN-hTERT</sup> clones and is consistent with the reactivation of telomerase, although this increase was not statistically significant ( $p=0.9591$ , Mann Whitney test). This increase in heterogeneity is reflected by the standard deviation values calculated at different time points. In contrast, the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones displayed consistent heterogeneous XpYp telomere length profiles, independent of the PD analysed. This is demonstrated by significantly greater standard deviation values obtained for the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones before, during and after crisis compared with the HAP1 WT<sup>DN-hTERT</sup> clones ( $p<0.05$ ,  $p<0.001$  and  $p<0.05$  respectively, Mann Whitney test) (**Figure 3.10D**).

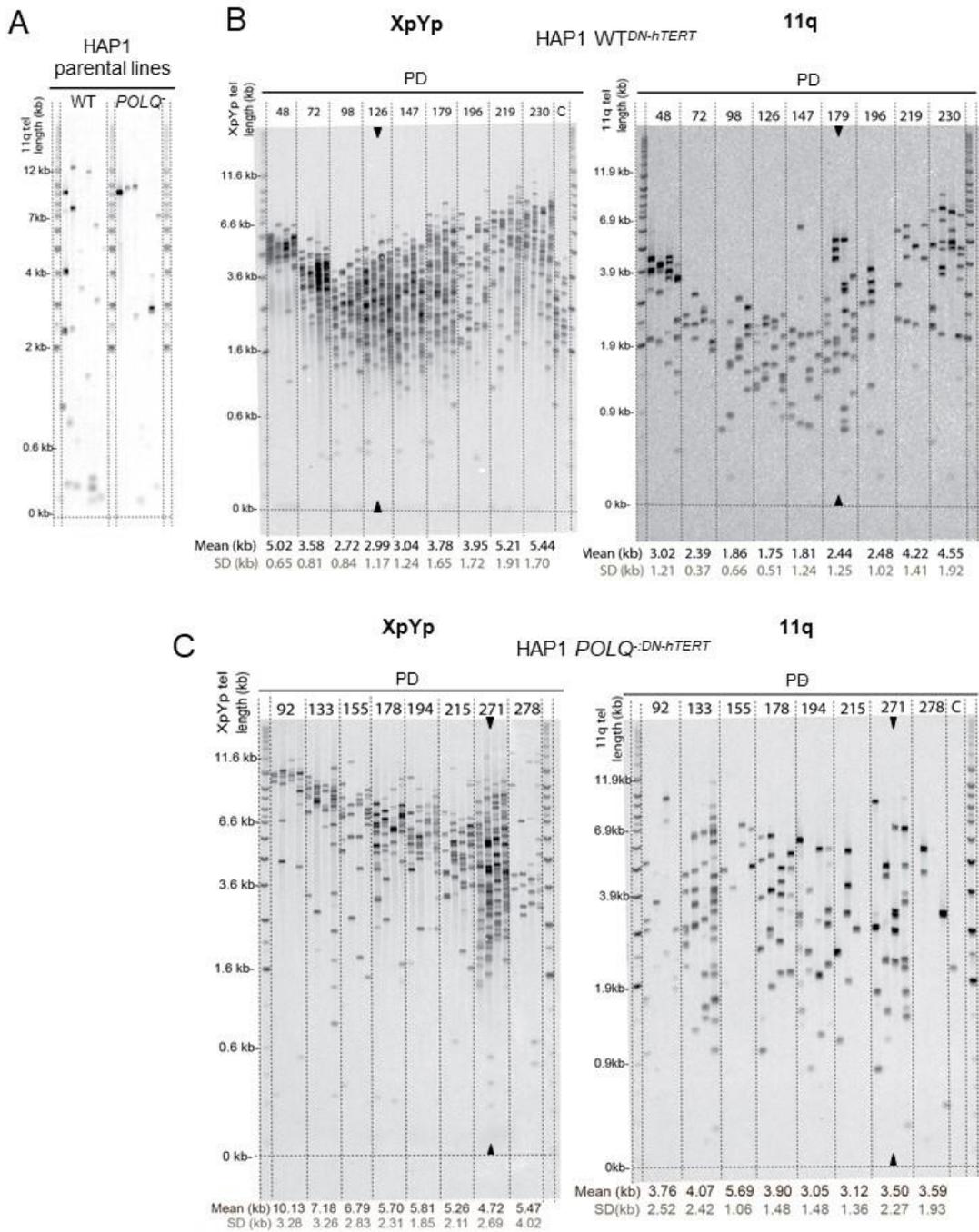


**Figure 3.10:** STELA profile for HAP1 WT<sup>DN-hTERT</sup> C11 (**A**) and HAP1 POLQ<sup>:DN-hTERT</sup> C2 at the XpYp telomere alongside a PCR control (C). The time point associated with telomere-driven crisis is highlighted with black arrows. (**B**) Scatter plot depicting mean shortest XpYp telomere lengths calculated at the point of crisis for HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> clones. The mean XpYp telomere length is displayed in red (WT) or black (POLQ) alongside standard deviation intervals. Significance was determined using a Mann Whitney test. (**C**) Graphical representation of STELA profiles obtained for HAP1<sup>PURO</sup> C1 at the XpYp telomere for two different time points. 6 reactions performed per sample. Mean telomere length is displayed in red with standard deviation intervals. (**D**) Graph displaying standard deviation values at early, crisis and late time points for HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> clones. The mean standard deviation is displayed in red (WT) or black (POLQ). Significance was determined using a Mann Whitney test.

### 11q telomere analysis

It is currently considered that telomere-driven crisis can be initiated by the accumulation of just a few critically short telomeres. Therefore, I decided to analyse telomeres at an additional chromosome end to determine whether the telomere dynamics observed at the XpYp telomere were consistent with the telomere dynamics at another chromosome end. As STELA of HAP1 WT<sup>DN-hTERT</sup> C3 (senescent clone) at the 11q telomere showed much shorter telomere lengths than the XpYp telomere, (Figure 3.8D, Section 3.4.7.1) I performed STELA for HAP1 WT<sup>DN-hTERT</sup> C1 (crisis clone) at the 11q telomere. This analysis showed telomere erosion of the 11q telomere to a length of 1.57 kb and subsequent telomere elongation to a length of 4.55 kb, similar to the telomere erosion and elongation observed at the XpYp telomere. However, the PD at which telomere elongation was detected (PD 179) was later than the XpYp telomere (PD 126). (Figure 3.11A-B). This was unexpected as the 11q telomere is shorter than the XpYp telomere in this clone and the current hypothesis is that shortest telomeres are preferentially elongated by telomerase (Britt-Compton et al. 2009).

Interestingly, the same pattern of telomere erosion and stabilisation was not observed at the 11q telomere for the HAP1 *POLQ*<sup>-</sup> clones. STELA of the 11q telomere in HAP1 *POLQ*<sup>-</sup>:DN-hTERT C2 showed an increase in heterogeneity of telomere lengths at each time point sampled (Figure 3.11C-D). This increased heterogeneity was also observed in seven additional clones (data not shown) suggesting that this observation is common amongst HAP1 *POLQ*<sup>-</sup>:DN-hTERT clones. STELA of the 11q telomere for HAP1 WT and *POLQ*<sup>-</sup> parental cell lines indicates that 11q telomere lengths may also be too long to amplify successfully using this PCR method.



**Figure 3.11: (A)** STELA profile for HAP1 WT and HAP1 *POLQ* parental cell lines at 11q telomere. 6 reactions performed per sample. **(B)** STELA profile of HAP1  $WT^{DN-hTERT}$  C1 at the XpYp telomere (left) and 11q telomere (right) at 9 time points (4 reactions per sample). Stabilisation of XpYp telomere length at PD 126 is highlighted in black. Stabilisation of 11q telomere length at PD 179 is highlighted by black arrows. **(C)** STELA profile of HAP1 *POLQ:DN-hTERT* C2 at the XpYp telomere (left) and 11q telomere (right) at 8 time points (4 reactions per sample). Stabilisation of XpYp and 11q telomere lengths at PD 179 is highlighted by black arrows.

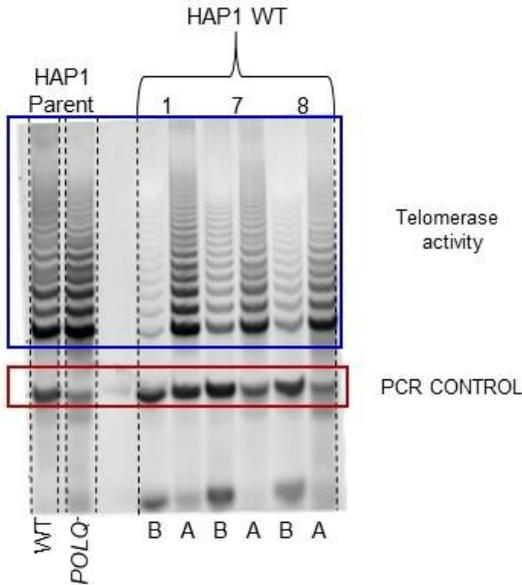
### TRAP Assay

It has been shown that cancer cells can achieve proliferative immortality via the upregulation or activation of telomerase. Furthermore, it has been demonstrated that telomerase is expressed in ~90% of human cancers (Yuan et al. 2019). An alternative, recombination dependent mechanism, alternative lengthening of telomeres (ALT) is responsible for telomere elongation in ~10% of human cancers (Dilley and Greenberg 2015).

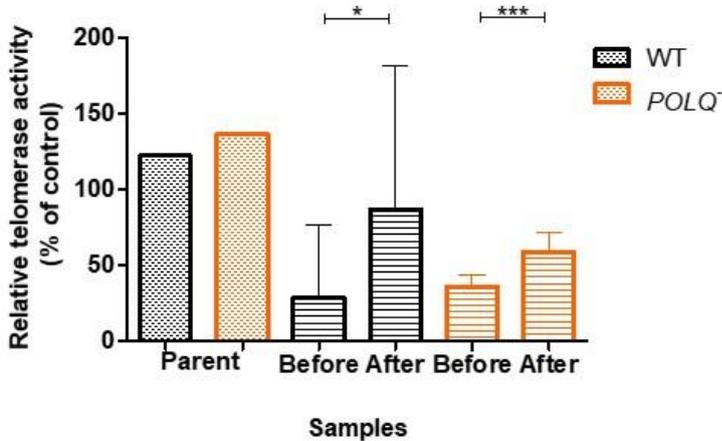
To determine the method of telomere length stabilisation in both HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-DN-hTERT</sup> clones following telomere-driven crisis, a TRAP assay was performed. HAP1<sup>DN-hTERT</sup> clones were analysed before and after crisis in addition to HAP1 WT and HAP1 POLQ<sup>-</sup> parental lines. This PCR based assay demonstrates the presence of telomerase activity via the generation of telomere repeats (Mender and Shay 2015). Telomeric repeats were separated by gel electrophoresis and visualised following staining of the gel with a nucleic acid gel stain, SYBR gold. The presence of telomeric DNA was quantified using ImageQuant. The ratio of intensity levels of TRAP sample ladders and the internal control (IC) band (intensity of sample's TRAP ladder/intensity of IC band) was calculated and each sample was normalised to the positive control as a percentage) after background subtraction.

Relative telomerase activity (as a % of the control) for the HAP1 WT cell line was 122% compared to 136% for the HAP1 POLQ<sup>-</sup> parental line. A downregulation of telomerase activity was observed in all HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-DN-hTERT</sup> clones before crisis. A 4.3-fold and 2.9-fold average decrease was calculated for HAP1 WT and POLQ<sup>-DN-hTERT</sup> clones before crisis compared to the parental lines. Following crisis, there was a 3-fold and 1.3-fold average increase in telomerase activity for HAP1 WT and POLQ<sup>-DN-hTERT</sup> clones respectively. The large standard deviation values calculated for these clones after crisis suggest that clones exhibited different levels of telomerase expression, with some clones demonstrating levels comparable those to the parental lines (**Figure 3.12**).

A



B



**Figure 3.12: (A)** TRAP assay blot with HAP1 parental lines (left) and HAP1 WT<sup>DN-hTERT</sup> clones before (B) and after (A) crisis. Bands representing telomeric DNA highlighted by blue box and PCR control highlighted in red box. **(B)** Bar chart showing relative telomerase activity (as a % of control) for HAP1 parental lines in addition to HAP1<sup>DN-hTERT</sup> clones (WT n=8, POLQ<sup>-</sup> n=10) before and after crisis. Error bars represent standard deviation values. Significance was determined using a Mann Whitney test.

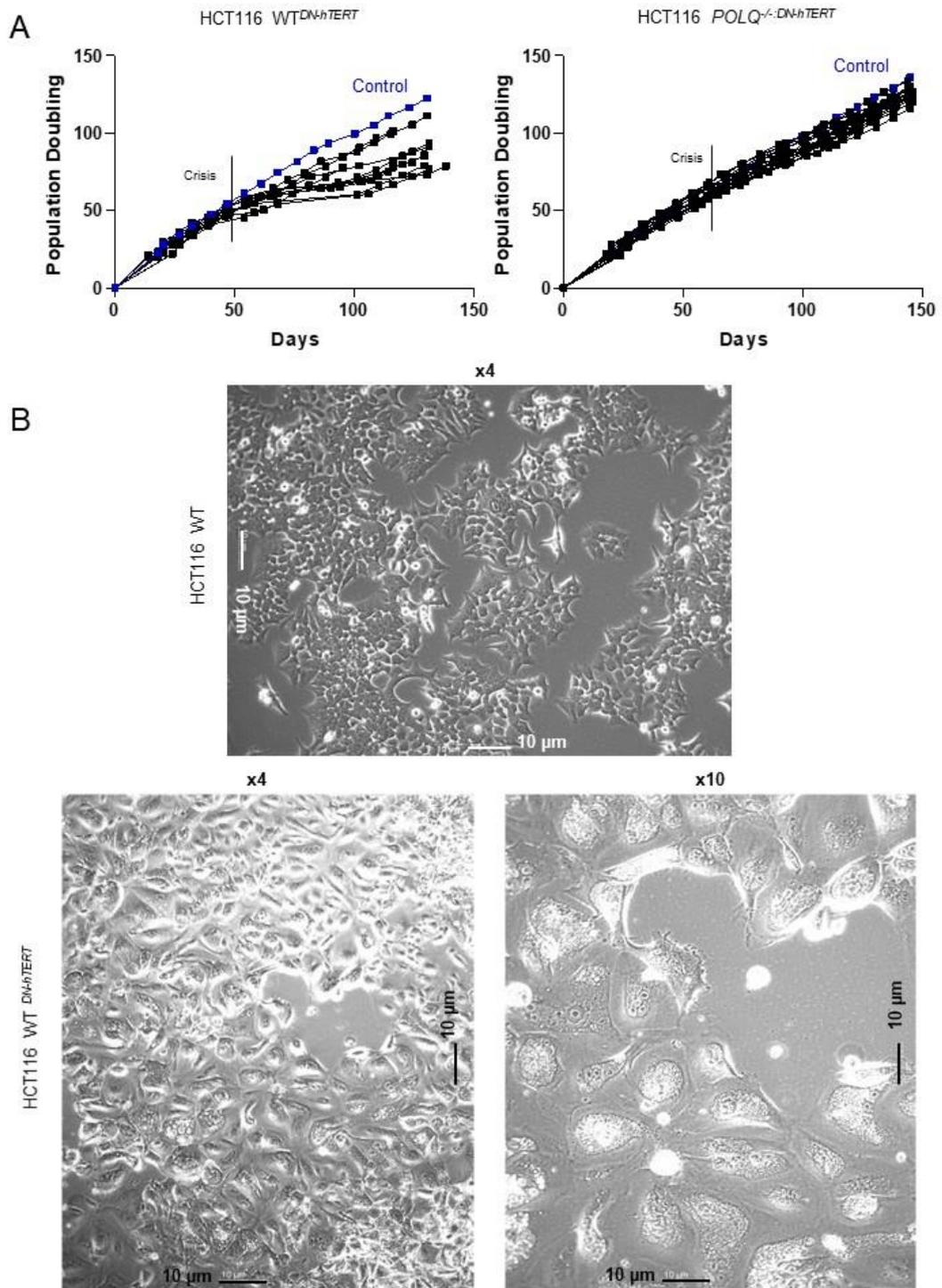
### 3.4.8 A Telomere-Driven Crisis in the HCT116 Cell Line

The HCT116 cancer cell line also maintains telomere length by telomerase activity rather than the alternative lengthening of telomeres (ALT) mechanism. To force these cells into a telomere-driven crisis, telomerase activity was inhibited by transduction with amphotropic retroviral vectors (pBABE-puro) containing a dominant-negative telomerase reverse transcriptase construct (DN-hTERT), as for the HAP1 cells. The pBABE-puro vector was used to generate control clones (HCT116 WT<sup>PURO</sup> and HCT116 POLQ<sup>-/-:PURO</sup>). Successful incorporation of the DN-hTERT or the empty pBABE-puro vector was selected for using puromycin 24 hr following transduction of HCT116 WT<sup>DN-hTERT</sup> and HCT116 POLQ<sup>-/-:DN-hTERT</sup> cells (day 1). Single cell clones were picked using cloning rings (11 WT<sup>DN-hTERT</sup>, 16 POLQ<sup>-/-:DN-hTERT</sup>, HCT116 WT<sup>PURO</sup> and HCT116 POLQ<sup>-/-:PURO</sup>) and population growth was monitored with regular DNA sampling for each clonal population up to 155 days following transduction. STELA was performed to compare telomere dynamics throughout the duration of the experiment. Of the 11 HCT116 WT<sup>DN-hTERT</sup> and 16 HCT116 POLQ<sup>-/-:DN-hTERT</sup> clones that were maintained in culture, there was a 100% survival rate. Likewise, HCT116 WT<sup>PURO</sup> and HCT116 POLQ<sup>-/-:PURO</sup> controls also survived the duration of the experiment.

#### Growth Curves

Following transduction with the DN-hTERT cassette, HCT116 WT<sup>DN-hTERT</sup> and HCT116 POLQ<sup>-/-:DN-hTERT</sup> clones were passaged in culture for up to 150 days, at which point clones were frozen and stored. During this time in culture, the HCT116 WT<sup>DN-hTERT</sup> clones completed between 75 and 111 PDs and the HCT116 POLQ<sup>-/-:DN-hTERT</sup> clones completed between 116 and 133 PDs. HCT116 WT and HCT116 POLQ<sup>-/-</sup> clones were passed on average every 7 days. A period of slowed population growth was observed for HCT116 WT<sup>DN-hTERT</sup> clones that was associated with enlarged cell size, increased cell death and the appearance of multi-nucleated cells (**Figure**

**3.13A).** These characteristics were used to define the onset of a telomere-driven crisis. Population growth was slowed during this time for a period between 12 and 46 days. The average number of days of retarded cell growth for the HCT116 WT<sup>DN-hTERT</sup> clones was 27. In contrast, retarded cell growth was not observed in the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones. Furthermore, no morphological indications of crisis were observed, and no apparent slowing of cell growth was associated with the onset of crisis in this cell line. Therefore, crisis was defined by the shortest telomere lengths observed using STELA and the presence of telomere fusion events. The control HCT116 WT and HCT116 *POLQ*<sup>-/-:PURO</sup> clones grew continuously for up to 150 days in culture without a change in growth kinetics (**Figure 3.13B**).



**Figure 3.13: (A)** Growth curves for HCT116 WT<sup>DN-hTERT</sup> (left) and HCT116 POLQ<sup>-/-</sup>DN-hTERT clones (right). HCT<sup>PURO</sup> controls highlighted in blue. **(B)** Microscope photograph of HCT116 WT parental cell line at x4 magnification (top) and HCT116 WT<sup>DN-hTERT</sup> C12 during crisis (passage 6) at 4x (bottom left) and 10x (bottom right) magnification. Scale bars shown in white (top) and black (bottom images).

### Telomere dynamics

STELA of the 17p telomere using the 17pseq1rev primer in clonal HCT116 cells expressing DN-hTERT produces bimodal telomere length profiles consistent with two allelic telomere length distributions at this chromosome end (Jones et al. 2014). STELA of the HCT116 WT<sup>DN-hTERT</sup> clones demonstrated telomere erosion of both alleles, with the shortest allele decreasing to a mean 17p telomere length of 0.85 kb, following which telomere length was elongated and maintained at a mean length of 2.00 kb (n=11). Once telomere lengths were maintained, it was not possible to discern between the two alleles, instead a single heterogeneous distribution is observed. Similarly, STELA of the HCT116 *POLQ*<sup>-/-DN-hTERT</sup> clones demonstrated telomere erosion of both alleles, with the shortest allele decreasing to a longer mean 17p telomere length of 1.65 kb ( $p>0.01$ , Mann Whitney test compared with WT, n=16) following which telomere lengths were maintained at an average 17p telomere length of 2.15 kb (**Figure 3.14A**).

### Starting telomere lengths

The mean starting 17p telomere length of the longer allele for the HCT116 WT<sup>DN-hTERT</sup> clones following transduction (~PD 20) was 4.53 kb, compared to the mean length of 1.86 kb for the shorter allele. Concurrent with the observation that the parental HCT116 *POLQ*<sup>-/-</sup> cell line had a longer 17p telomere length than the HCT116 WT cell line, the mean starting 17p telomere length (~PD 20) of the longer allele for the HCT116 *POLQ*<sup>-/-DN-hTERT</sup> clones was 6.22 kb ( $p>0.01$ , Mann Whitney test compared with WT). The mean length of the shorter allele for the HCT116 *POLQ*<sup>-/-DN-hTERT</sup> clones was 3.64 kb ( $p>0.01$ , Mann Whitney test compared with WT). For clones that had a STELA profile in which the two alleles were too similar in length to be discerned, the mean 17p starting telomere length for the HCT116 WT<sup>DN-hTERT</sup> clones was 2.41 kb

compared with 4.31 kb for the HCT116 *POLQ*<sup>-/-DN-hTERT</sup> clones ( $p > 0.01$ , Mann Whitney test) (**Figure 3.14B**).

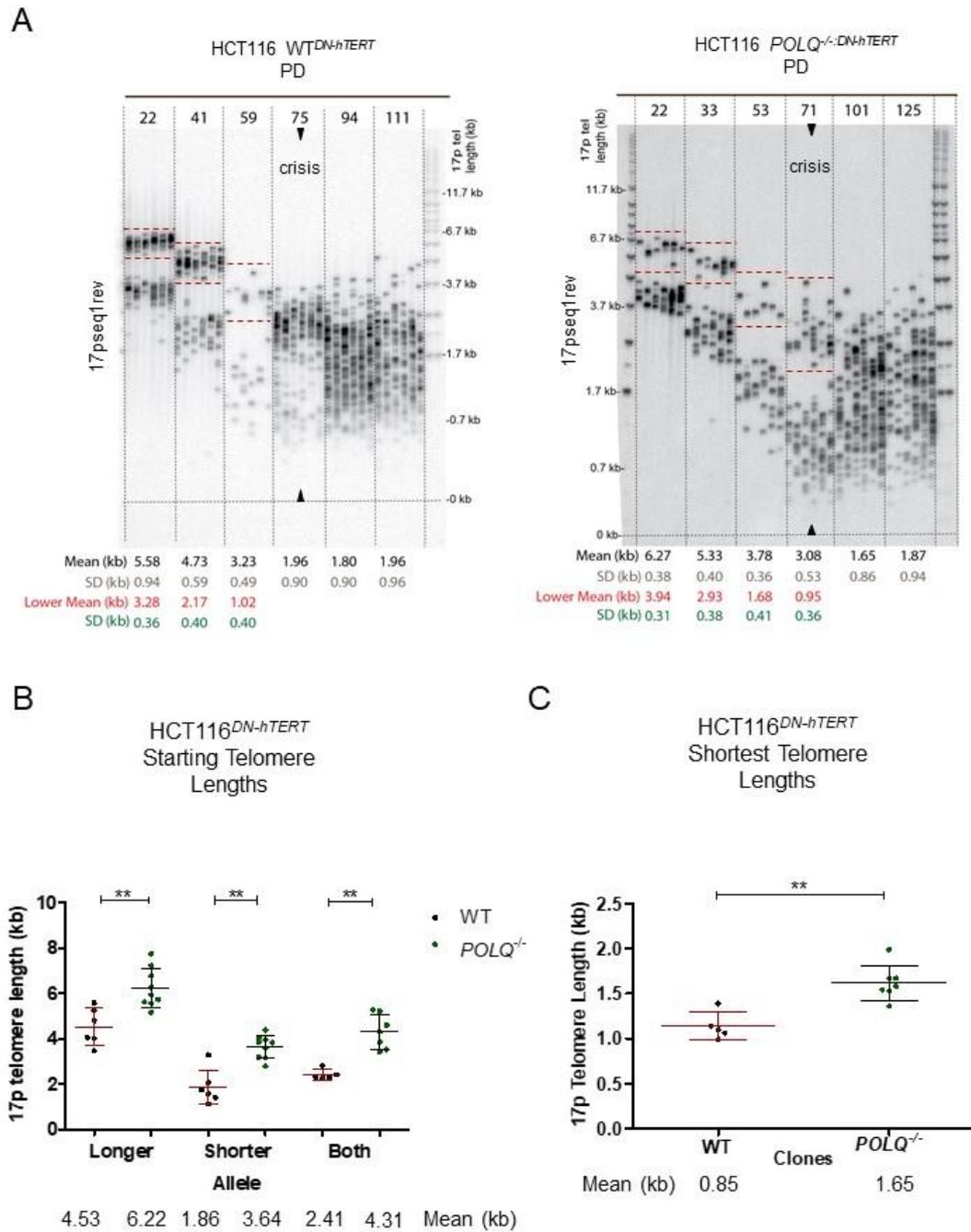
There was no statistically significant difference in the rate of telomere erosion when comparing the two cell lines. The HCT116 WT<sup>DN-hTERT</sup> cell line had a mean 17p telomere erosion rate of 45 bp per PD compared to a mean 17p telomere erosion rate of 51 bp per PD for the HCT116 *POLQ*<sup>-/-DN-hTERT</sup> cell line ( $p = 0.1339$ , Mann Whitney test). HCT116 WT<sup>DN-hTERT</sup> clones entered a period of telomere-driven crisis at shorter 17p telomere lengths than the HCT116 *POLQ*<sup>-/-DN-hTERT</sup> clones (**Figure 3.14C**). To determine whether these telomere dynamics were replicated at additional chromosome ends, STELA was also performed at the XpYp telomere for HCT116 WT<sup>DN-hTERT</sup> C21 and HCT116 *POLQ*<sup>-/-DN-hTERT</sup> C5. The same pattern of telomere erosion and stabilisation was observed at the XpYp telomere for both HCT116 WT C21 and HCT116 *POLQ*<sup>-/-DN-hTERT</sup> C5, although this telomere only has one allelic distribution (**Figure 3.14D**).

In contrast, STELA of HCT116 WT<sup>PURO</sup> and HCT116 *POLQ*<sup>-/-PUR0</sup> control clones at an early (PD 20/23), mid (PD 70/75) and late (PD 153) time point showed no change in the mean 17p telomere length. As observed with the HCT116 parental lines, the HCT116 WT<sup>PURO</sup> mean 17p telomere lengths calculated for these time points (3.26 kb, 2.98 kb and 2.92 kb) were significantly shorter than the mean 17p telomere lengths calculated for the HCT116 *POLQ*<sup>-/-PUR0</sup> clones (4.77 kb, 4.85 kb and 4.21 kb) ( $p > 0.001$ , Mann Whitney test) (**Figure 3.14E**).

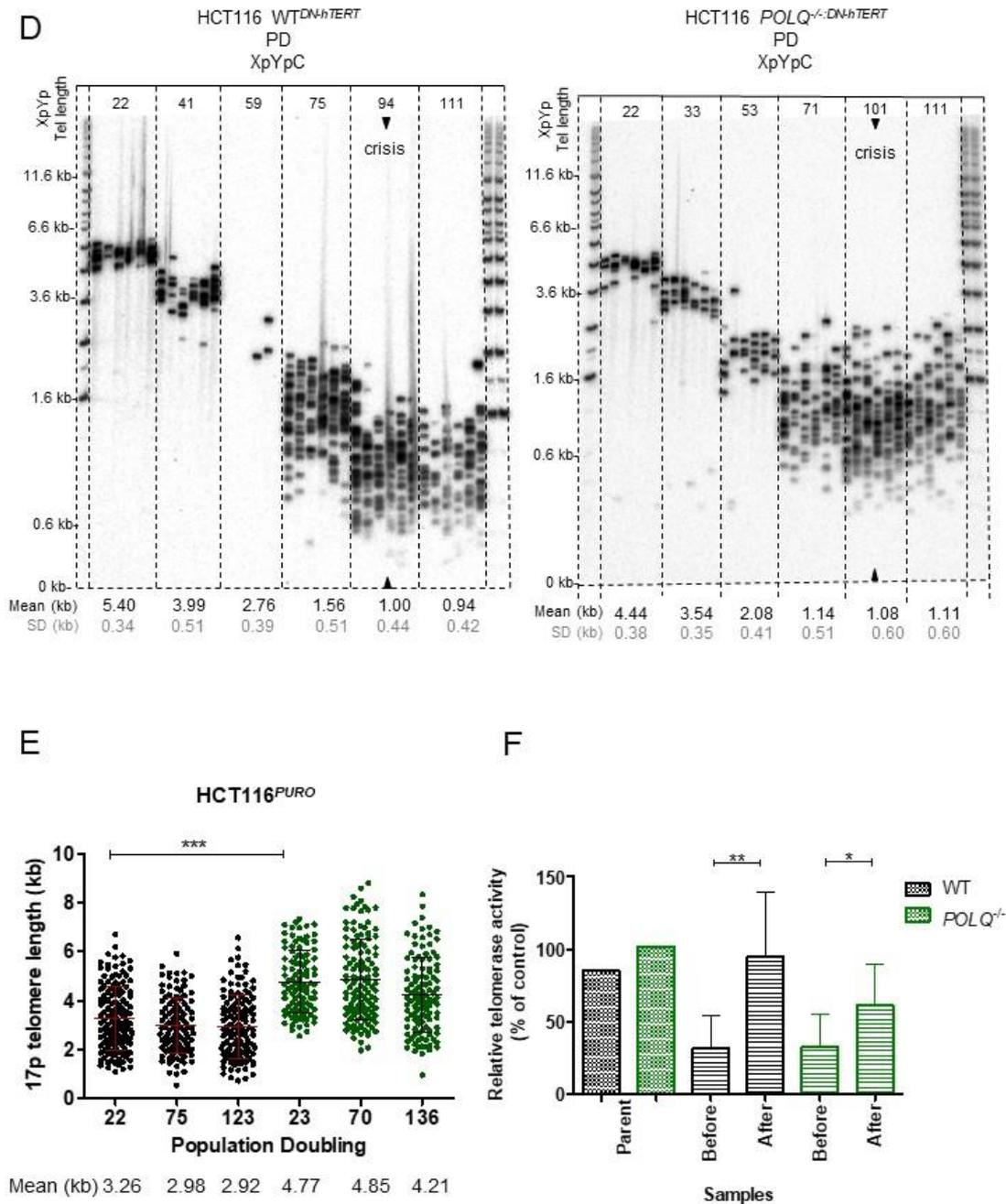
### TRAP Assay

To determine whether telomerase activity was responsible for the stabilisation of telomere length following crisis, a TRAP assay was performed. HCT116<sup>DN-hTERT</sup> clones were sampled before and after crisis in addition to HCT116 parental lines. Relative telomerase activity (as a % of the control) for the HCT116 WT cell line was

85% compared to 102% for the HAP1 *POLQ*<sup>-</sup> parental line. A downregulation of telomerase activity was observed in all HCT116<sup>DN-hTERT</sup> clones prior to crisis as a 2.7-fold and 3.2-fold average decrease was calculated for HCT116 WT<sup>DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-DN-hTERT</sup> clones in comparison to parental lines. An upregulation in telomerase activity compared to pre-crisis samples was observed in all post-crisis HCT116<sup>DN-hTERT</sup> clones as a 3-fold and 1.9-fold average increase was calculated for HCT116 WT<sup>DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-DN-hTERT</sup> clones, respectively. **(Figure 3.14F).**



**Figure 3.14: (A)** STELA profiles for HCT116 WT<sup>DN-hTERT</sup> (left) and HCT116 POLQ<sup>-/-</sup>:DN-hTERT clones (right) at the 17p telomere. Telomere-driven crisis is annotated and highlighted by black arrows. Longer allele highlighted by red dashed lines. **(B)** Chart showing mean starting 17p telomere lengths for HCT116 WT<sup>DN-hTERT</sup> and HCT116 POLQ<sup>-/-</sup>:DN-hTERT clones at both alleles. **(C)** Chart showing mean shortest 17p telomere lengths observed for HCT116 WT<sup>DN-hTERT</sup> and HCT116 POLQ<sup>-/-</sup>:DN-hTERT clones. Mean is shown in red (WT) and black (POLQ<sup>-/-</sup>) alongside standard deviation intervals. Significance was determined using a Mann Whitney test.



**Figure 3.14:**(D) STELA profiles for HCT116 WT<sup>DN-hTERT</sup> (left) and HCT116 POLQ<sup>-/-</sup>DN-hTERT (right) clones at the XpYp telomere. Telomere-driven crisis is annotated and highlighted by black arrows. (E) Graphical representation of 17p STELA profiles for HCT116 WT<sup>PURO</sup> and HCT116 POLQ<sup>-/-</sup>PURO clones at an early, mid and late time point. 6 reactions performed per sample. Mean telomere length is highlighted in red (WT) or black (POLQ<sup>-/-</sup>) alongside standard deviation intervals. Significance was determined using a Mann Whitney test. (F) Bar chart showing relative telomerase activity (as a % of control) for HCT116 parental lines in addition to HCT116<sup>DN-hTERT</sup> clones (WT n=11, POLQ<sup>-/-</sup> n=16) before and after crisis. Error bars represent standard deviation values. Significance was determined using a Mann Whitney test.

### 3.5 DISCUSSION

#### 3.5.1 Characteristics of POLQ-deficient Cell Lines

The disruption of *POLQ* in three distinct cell lines, HAP1, HCT116 and RPE1<sup>hTERT</sup>, was achieved using CRISPR/Cas gene editing. Disruption of *POLQ* using this method did not have a significant effect on cell viability or growth rates in any of the cell lines. As the *POLQ*-deficient cell lines were not from the same populations as the WT, this could explain the small differences observed in cell growth for the HAP1 and RPE1<sup>hTERT</sup> cell lines. These cell lines may have spent differing amounts of time in culture and experienced differing levels of cellular stress. As *POLQ* is involved in the A-NHEJ pathway utilised following DNA damage (Yousefzadeh et al. 2014), it is conceivable that disruption of this gene would not cause changes in cellular phenotype until cells experience DNA damage.

Initially, STELA was used to determine telomere lengths at specific chromosome ends in the parental cell lines. Interestingly, the *POLQ*-deficient cell lines displayed significantly longer XpYp telomere lengths than the WT in the three distinct cell lines (HAP1  $p < 0.01$ , HCT116  $p < 0.001$ , RPE1<sup>hTERT</sup>  $p < 0.001$ , Mann Whitney test). This significant difference in length was observed at three distinct chromosome ends (XpYp, 17p, 5p) for the RPE1<sup>hTERT</sup> cell lines. However, as the HAP1 *POLQ*<sup>-</sup> cell line had very long telomeres, STELA proved to be problematic due to the constraints of amplifying such large molecules via PCR. Therefore, a telomere restriction fragment (TRF) assay was performed to determine the mean telomere length of the parental cell lines across all chromosome ends. This assay demonstrated that *POLQ*-deficient cell lines had longer telomere lengths than the WT cell lines for all three parental cell lines.

Whilst it is an interesting observation that telomere lengths are longer in three distinct *POLQ*-deficient cell lines, it is not possible to determine whether this difference in telomere length is due to the disruption of *POLQ* as the *POLQ*-deficient cell lines are

not from the same parental populations as the WT for any of the cell lines. It is possible that the disruption of *POLQ* via CRISPR/Cas gene editing impacts cellular survival. Therefore, only cells that had long telomeres were able to survive the gene editing process. Further work is needed to clarify whether disruption of the *POLQ* gene affects telomere length. This work would involve disruption of *POLQ* in clonal populations of WT cell lines so a direct comparison can be made between the disruption of the *POLQ* gene and telomere length. However, the elongation associated with the *POLQ*-deficient cell lines, particularly the HAP1 cell line, are beyond the range of the telomere lengths observed in the WT cell lines. Alternatively, *POLQ*-deficient cells could be complemented with WT *POLQ* to see if this would rescue the phenotype, i.e. that shorter telomere lengths are restored. This could be repeated with mutated versions of *POLQ* to identify the specific domains that are required for the telomere elongation activity.

If the disruption of *POLQ* does lead to telomere elongation it could be due to the interaction of the helicase domain of *POLQ* with RPA at telomeric ssDNA. It has previously been demonstrated that POT1 antagonises RPA at telomeric DNA (Barrientos et al. 2008; Flynn et al. 2011), but how POT1 executes this role is not understood. In addition, studies have shown that RPA is able to bind to telomeric ssDNA as or more efficiently than the POT-TPP1 complex *in vitro* (Barrientos et al. 2008; Takai et al. 2011). It may be that *POLQ* is able to associate with POT1 and facilitate the removal of RPA from telomeric ssDNA as it has been shown that the helicase domain is able to facilitate the removal of RPA from resected DSBs (Mateos-Gomez et al. 2017). Furthermore, RPA transiently associates with telomeres during S-phase of the cell cycle (Mcgee et al. 2010) and RPA is also clearly implicated in telomere maintenance in yeast (Smith et al. 2000). Therefore, the disruption of *POLQ* may result in the accumulation of RPA at telomeric ssDNA which prevents the binding

of POT1, allowing telomerase to access the 3' single-stranded overhang and results in the upregulation of telomere maintenance mechanisms.

All three WT and POLQ-deficient lines were telomerase positive cell lines. To determine whether telomerase is required for the elongation of telomeres following the disruption of *POLQ*, the disruption of *POLQ* in cell lines which do not express telomerase could be performed.

### 3.5.2 Telomere-driven crisis in HAP1 and HCT116 cell lines

#### Clonal population growth

To assess the impact of POLQ on the ability of cells to escape telomere-driven crisis, HAP1 WT, HAP1 *POLQ*<sup>-/-</sup>, HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cells were transduced with a DN-hTERT cassette to induce telomere erosion.

Clonal populations of both HAP1 *POLQ*<sup>-/-DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-DN-hTERT</sup> escaped crisis more readily than the HAP1 and HCT116 WT cell lines. In addition to more stable population growth during crisis, no morphological changes were linked to the onset of telomere-driven crisis in either HAP1<sup>DN-hTERT</sup> or HCT116<sup>DN-hTERT</sup> *POLQ*-deficient cell lines whereas increased cell size and the presence of multinucleated cells associated with telomere crisis were observed for both HAP1<sup>DN-hTERT</sup> and HCT116<sup>DN-hTERT</sup> WT cell lines.

These data could indicate that DNA repair mechanisms utilised during crisis have an impact on the cellular capacity for escape. As POLQ is a component of the A-NHEJ repair pathway, these data may suggest that the inhibition of intra-chromosomal fusion associated with A-NHEJ can precipitate recombinations that result in telomerase being reactivated more quickly, therefore allowing cells to escape crisis. As A-NHEJ is inhibited in these cell lines, it is likely that these telomere fusion events are products of the C-NHEJ pathway which are associated with inter-chromosomal fusion events. However, this does not reflect previously published data which

suggests that increased levels of intra-chromosomal fusion events mediated by A-NHEJ facilitate the escape from crisis (Jones et al. 2014). In addition to the activity of the polymerase domain of POLQ, the helicase domain of POLQ may disrupt DSB repair mechanisms such as HR. The disruption of POLQ could therefore result in the upregulation of alternative DSB repair mechanisms such as HR and C-NHEJ and cause an increase in inter-chromosomal fusion events. To further examine the role of these repair pathways, fusion PCR was performed on HAP1 WT<sup>DN-hTERT</sup>, HAP1 POLQ<sup>-DN-hTERT</sup> as well as HCT116 WT<sup>DN-hTERT</sup> and HCT116 POLQ<sup>-/-DN-hTERT</sup> clones as they progressed through crisis (*Chapter 4*).

### Telomere length dynamics

As observed with the HAP1 parental cell lines, the starting XpYp telomere lengths for HAP1 WT<sup>DN-hTERT</sup> clones were significantly shorter than the XpYp starting telomere lengths for HAP1 POLQ<sup>-DN-hTERT</sup> clones. Of the 10 HAP1 WT<sup>DN-hTERT</sup> clones, the two clones that entered replicative senescence had the shortest starting XpYp telomere lengths of 4.42 kb and 5.92 kb, with one exception, C12 which had a starting XpYp telomere length of 5.54 kb. This suggests that these clones had less of a telomere buffer permissive of the genomic rearrangements required for crisis escape.

Both HAP1 POLQ<sup>-DN-hTERT</sup> and HCT116 POLQ<sup>-/-DN-hTERT</sup> cell lines had longer telomere lengths than the HAP1 WT<sup>DN-hTERT</sup> and HCT116 WT<sup>DN-hTERT</sup> cell lines at the point of crisis ( $p > 0.001$ , Mann Whitney test). Whilst the shortest XpYp telomere length observed for the HAP1 WT<sup>DN-hTERT</sup> clones was consistent, the shortest XpYp telomere lengths detected amongst the HAP1 POLQ<sup>-DN-hTERT</sup> clones were highly variable. The current hypothesis is that it only takes the accumulation of a few critically short telomeres to initiate telomere-driven crisis (Counter et al. 1992). Thus, it could be that telomeres at other chromosomes have become critically short in these clones and initiated telomere-driven crisis. The XpYp telomeres in these HAP1 POLQ<sup>-DN-hTERT</sup> clones could be longer and more variable than telomeres at different chromosome

ends. However, it may be possible that the absence of POLQ renders telomeres more prone to uncapping resulting in the initiation of a DDR at longer telomere lengths.

Unlike the HAP1 *POLQ*<sup>-DN-hTERT</sup> clones, the (17p) telomere length at which crisis was induced in the HCT116 cell line was consistent amongst both the HCT116 WT and the *POLQ*<sup>-/-:DN-hTERT</sup> clones.

### 11q Telomere Length Analysis

STELA of HAP1 WT<sup>DN-hTERT</sup> clones at the 11q telomere shows telomere erosion induced by the abrogation of telomerase followed by stabilisation of telomere length as observed at the XpYp telomere. However, the XpYp telomere length stabilised earlier (PD 126) than the 11q telomere length (PD 179). This is unexpected as the 11q telomere is shorter than the XpYp telomere and the current hypothesis is that shortest telomeres are preferentially elongated by telomerase (Britt-Compton et al. 2009). This delay in telomere elongation may be due to the structure of the 11q telomere. If the structure of this telomere is less accessible to telomerase, there could be a delay in telomerase access and therefore telomere length stabilisation.

In contrast, STELA of the 11q telomere in HAP1 *POLQ*<sup>-DN-hTERT</sup> clones showed extremely heterogeneous telomere distributions. This increase in telomere length heterogeneity was observed in seven additional HAP1 *POLQ*<sup>-DN-hTERT</sup> clones. STELA of the HAP1 *POLQ* parental cell line suggests long telomeres which cannot be effectively amplified using this PCR method. It is therefore not possible to determine whether the heterogeneous telomere length profiles observed at the 11q telomere was a result of the progression through and escape from crisis or a feature of the parental line.

### 3.5.3 Conclusions

Taken together, these data suggest that the disruption of *POLQ* may have an impact on telomere maintenance mechanisms, demonstrated by increased telomere lengths for *POLQ*-deficient cell lines compared with WT cell lines. Furthermore, the disruption of *POLQ* is associated with a greater cellular capacity for crisis escape. This is likely due to a change in the DSB repair mechanisms employed during crisis and therefore the formation of different types of telomere fusion events. As *POLQ* is a component of the A-NHEJ pathway it is likely that disruption of *POLQ* caused an inhibition of A-NHEJ in these cells and therefore a reduction in the frequency of intra-chromosomal fusion events. Previous data suggest that the escape from crisis is facilitated by A-NHEJ (Jones et al. 2014), so telomere fusion analysis of HAP1<sup>DN-hTERT</sup> and HCT116<sup>DN-hTERT</sup> cell lines will provide insight into whether the disruption of *POLQ* caused any changes in the abundance or type of fusion events that occurred during crisis (Chapter 4).

## Chapter 4

### What is the role of POLQ in the NHEJ pathways utilised for telomere fusion?

#### 4.1 ABSTRACT

To determine whether POLQ has a role in telomere fusion events, fusion analysis was performed on HAP1 and HCT116 WT and POLQ-deficient cell lines transitioning through telomere-driven crisis. A greater abundance of 16p-21q inter-chromosomal fusion events was observed for HAP1 *POLQ<sup>:DN-hTERT</sup>* clones compared with HAP1 WT<sup>*DN-hTERT*</sup> clones. Similarly, a greater frequency of 17p-21q inter-chromosomal fusion events was observed for the HCT116 *POLQ<sup>-/-DN-hTERT</sup>* clones compared with HCT116 WT<sup>*DN-hTERT*</sup> clones transitioning through telomere-driven crisis. Furthermore, a significant reduction in the abundance of 17p:17p sister-chromatid fusion events was also detected for HCT116 *POLQ<sup>-/-DN-hTERT</sup>* clones in comparison with HCT116 WT<sup>*DN-hTERT*</sup> clones.

Telomere fusion analysis was also performed on HAP1, HCT116 and RPE1<sup>*hTERT*</sup> WT and POLQ-deficient cell lines that had undergone nucleofection with 17p and/or 16p/21q TALEN plasmid pairs to induce subtelomeric DSBs. No significant differences in the frequencies of inter-chromosomal fusion events were observed following these DSBs, suggesting that a common repair mechanism was utilised in both the WT and POLQ-deficient cell lines. However, a significantly greater proportion of 17p:17p intra-chromosomal fusion events was observed for the HCT116 WT cell line compared with the HCT116 POLQ-deficient cell line following nucleofection with 17p TALEN pairs. These experiments were also performed in a HCT116 *POLQ<sup>-/-LIG4<sup>-/-</sup></sup>* cell line which is both A-NHEJ and C-NHEJ-deficient. Reductions in the abundance of all kinds of telomere fusion events were observed for this cell line compared with both HCT116 WT and HCT116 *POLQ<sup>-/-</sup>* cell lines.

## 4.2 INTRODUCTION

The inappropriate repair of telomeric DNA during telomere-driven crisis (Capper et al. 2007) is mediated by DNA repair pathways such as C-NHEJ and A-NHEJ which facilitate telomere fusion events that give rise to genomic instability. Recombinations may occur which result in the reactivation of telomerase, such as mutations within the *hTERT* locus, which allows the cell to escape crisis (Halvorsen et al. 1999). The capacity for cells to escape crisis may be influenced by the type of telomere fusion events that occur during crisis. A-NHEJ facilitates intra-chromosomal telomere fusion whereas C-NHEJ is associated with inter-chromosomal telomere fusion events during telomere-driven crisis (Jones et al. 2014). In addition, this study showed that cells which failed to escape crisis had increased proportions of C-NHEJ-mediated inter-chromosomal fusion events.

A-NHEJ relies on 5'–3' resection of DNA by MRN-CtIP which reveals microhomologies that allow base pairing of the complementary strands. This drives annealing of the resected ends and promotes synapsis of the opposite ends of the DSB. The annealed intermediates are stabilised by POLQ which performs fill-in synthesis. Whilst the deletions at break sites are caused by nucleolytic processing, insertions arise following POLQ activity (Kent et al. 2015). Ligation of ends is catalysed by DNA ligase 3 (LIG3) or DNA ligase 1 (LIG1).

Whilst previous data has shown that POLQ is involved in repair at deprotected telomeres that are not protected by shelterin components (Mateos-Gomez et al. 2015), no previous study has investigated the role of POLQ during telomere-driven crisis in human cell lines.

### 4.3 AIMS OF THE CHAPTER

The purpose of this chapter was to determine whether POLQ has a role in telomere fusion events that arise during telomere-driven crisis or following DSBs induced in subtelomeric DNA by nucleofection with TALEN pairs. Telomere fusion events were detected using telomere fusion PCR.

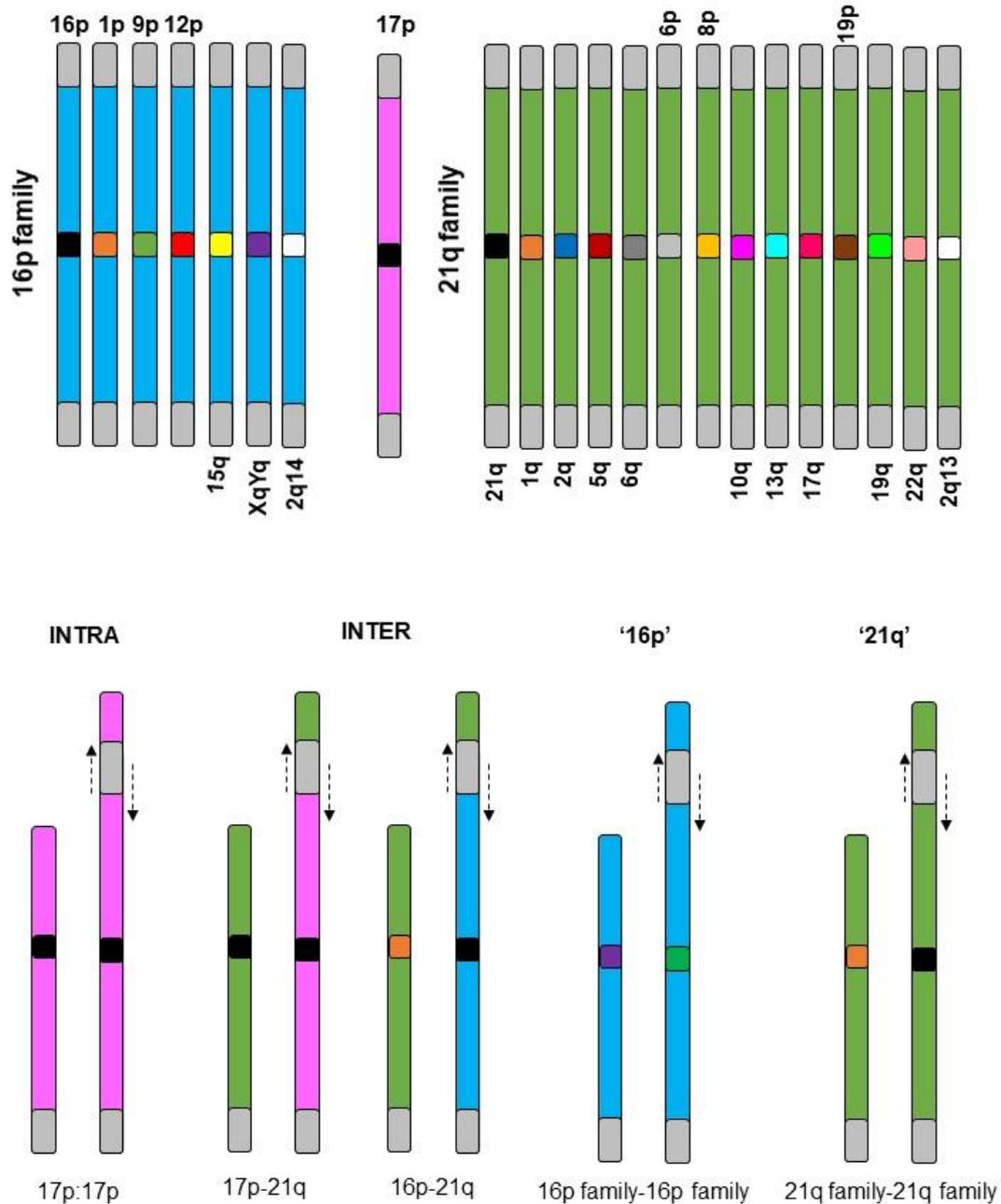
The aims of this chapter were as follows:

- To determine whether similar fusion profiles were observed for HAP1 <sup>DN-hTERT</sup> / HCT116 <sup>DN-hTERT</sup> WT and POLQ-deficient cells as they progressed through telomere-driven crisis
- To assess whether similar repair mechanisms were employed in HAP1, RPE1 <sup>hTERT</sup> and HCT116 WT and POLQ-deficient cell lines after specific subtelomeric DSBs were induced by nucleofection with 17p and 16p/21q TALEN pairs

#### 4.4 RESULTS

The onset of a telomere-driven crisis is characterised by increased genomic instability demonstrated by a rise in telomere fusion events. These end-to-end telomere fusions can be detected using telomere fusion PCR. Fusions between different chromosome ends or sister chromatids were detected by multi-primer, long-range single-molecule fusion PCR using single orientation primers, as described previously (*Chapter 2*). Following gel electrophoresis and Southern blotting, the use of chromosome-specific probes for hybridisation reveals which telomeres are involved in the fusion events. In order to establish any differences in the abundance or type (intra- or inter-chromosomal events) of telomere fusion events that arose during telomere-driven crisis, telomere fusion analysis was performed using genomic DNA extracted from HAP1<sup>DN-hTERT</sup> clones passaging crisis (*Chapter 3*).

Since the 17p primers were not able to amplify the 17p telomeres in the HAP1 cell line (*Chapter 3*), dual-primer analysis was performed using 16p and 21q primers instead of using 17p single-primer and 17p-21q dual-primer analysis. As the 16p and 21q primers target families of (rather than unique) telomeres, it is not possible to determine whether fusions amplified using the 16p or 21q primers are inter- or intra-chromosomal fusion events. These events are, therefore, termed '16p' or '21q' events in this chapter, rather than being described as inter- or intra-chromosomal fusions. Fusion events of the same molecular weight that are detected with both 16p and 21q subtelomere probes most likely reflect 16p-21q inter-chromosomal events (**Figure 4.1**).



**Figure 4.1:** (Top) The 16p and 21q fusion primers target families of telomeres (16p family shown in blue, 21q family shown in green) rather than unique telomeres due to shared sequence homology whereas the 17p primer targets the 17p telomere only. (Bottom) Telomere fusion events detected with the 17p primer only therefore represent 17p:17p intra-chromosomal fusion events (pink). Telomere fusion events detected between the 21q family (green) and the 17p telomere represent 17p-21q inter-chromosomal fusion events as do telomere fusion events between the 16p family (blue) and the 21q family. Telomere fusion events between 16p (16p-16p) or 21q family members (21q-21q) are termed '16p' or '21q' events as it is not possible to determine whether these events are inter- (9p-16p or 2q-5q) or intra-chromosomal (12p-12p or 10q-10q) fusion events with the primers used.

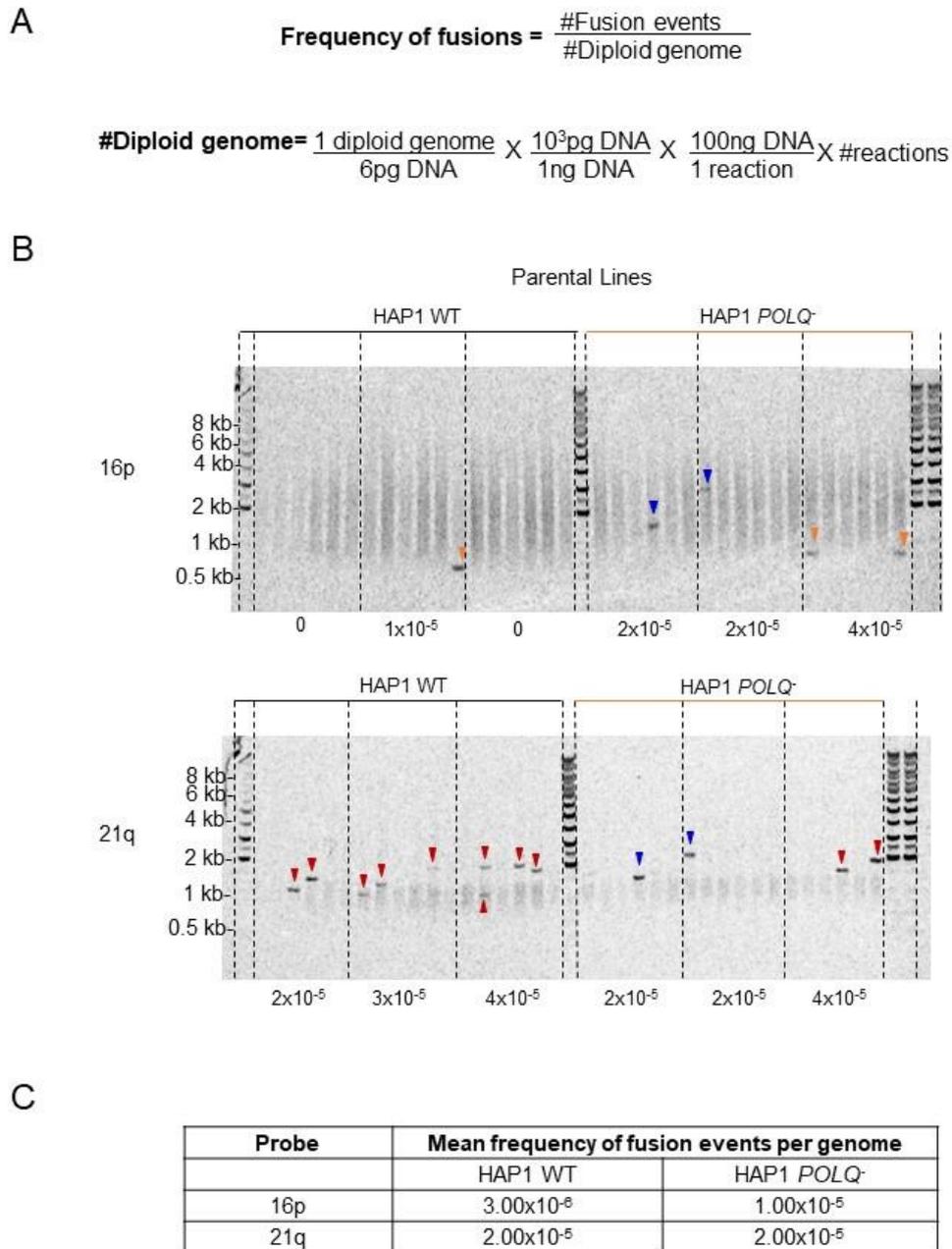
#### 4.4.1 Calculating the Frequency of Fusion Events

To determine whether there were any differences in the abundance of telomere fusion events, an equation was used to calculate the frequency of fusion events (**Figure 4.2A**). The value obtained from this calculation is a relative representation as this equation underestimates the frequency of genome-wide fusions. 25-100 ng DNA were used as input per fusion reaction depending on DNA availability. The calculation is based on an estimate of 6 pg DNA per diploid genome or 3 pg DNA per haploid genome. This calculation did not include any constitutive background bands (non-fusion amplicons). The ploidy of HAP1 samples was determined by image cytometry (*Chapter 3*). For the HAP1 *POLQ*<sup>-</sup> parental cell line (pre-transduction with DN-hTERT), the calculation assumed a haploid genome, however, as the HAP1 WT parental cell line was a mixture of haploid and diploid cells, the calculation for this cell line assumed a diploid genome. Similarly, as HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>-DN-hTERT</sup> clones became diploid early after transduction with the DN-hTERT cassette, the calculation made for these clones assumed a diploid genome.

For the parental lines, the mean frequency of fusion events was calculated at a single time point whereas for HAP1<sup>DN-hTERT</sup> clones, the maximal fusion frequency was calculated, i.e. the frequency of fusion events calculated for the crisis-transition time point at which most telomere fusion events were identified.

#### Fusion analysis of parental lines

Fusion analysis using 16p and 21q dual-primer PCR was performed on the parental lines to compare any increases in telomere fusion events that arose during telomere crisis. For the HAP1 WT parental cell line, the background frequency of fusion events calculated was  $3.00 \times 10^{-6}$  per diploid genome detected with the 16p probe compared to a frequency of  $3.00 \times 10^{-5}$  per diploid genome (p/dg) detected with the 21q probe. For the HAP1 *POLQ*<sup>-</sup> parental cell line, the frequency of fusion events detected with both 16p and 21q probes was  $2.70 \times 10^{-5}$  per haploid genome (p/hg) (**Figure 4.2B-C**).



**Figure 4.2:** (A) Equation used to calculate the frequency of fusion events. #=number (B) Southern blots of 16p:21q telomere fusion profiles hybridised with 16p probe (top) and 21q blot (bottom). 100 ng gDNA used per reaction which were repeated in triplicate. Fusion frequencies per genome shown beneath each sample calculated using the equation in (A). Fusion bands unique to 16p probe highlighted by orange arrows, fusion bands unique to the 21q probe highlighted by red arrows and fusion bands common to both probes highlighted by blue arrows. (C) Table showing the mean frequency of fusion events per genome for HAP1 WT and HAP1 *POLQ*<sup>-</sup> parental cell lines calculated using triplicate fusion frequency values calculated in (B).

#### 4.4.2 Fusion analysis of clones that did not reach telomere-driven crisis

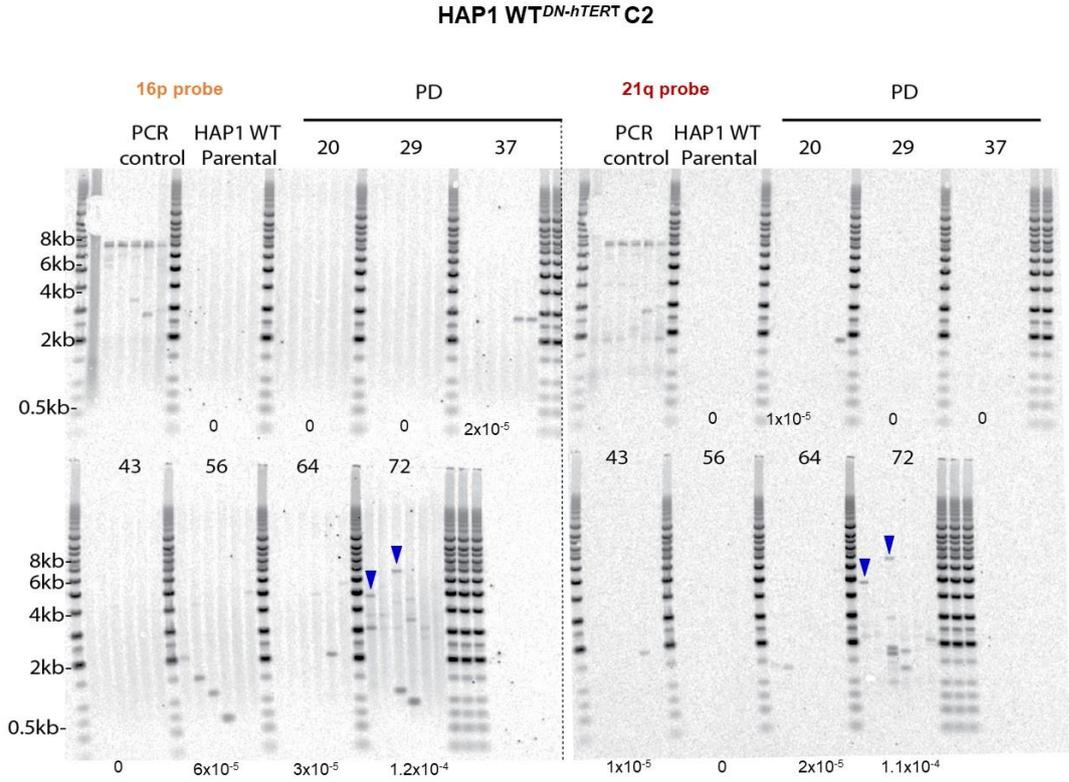
Two HAP1 WT<sup>DN-hTERT</sup> clones entered a period of static cell growth consistent with senescence. As these clones had shorter starting telomere lengths than other WT clones, it is possible that these clones had less of a telomere buffer to enable genomic rearrangements required for crisis escape (*Chapter 3*). Therefore, the maximal fusion frequencies of these clones following 16p:21q dual-primer PCR was calculated.

For HAP1 WT<sup>DN-hTERT</sup> Clone 2 (C2), the maximal fusion frequency of  $1.24 \times 10^{-4}$  p/dg was detected with the 16p probe compared with  $1.10 \times 10^{-4}$  p/dg for fusion events observed with the 21q probe at the same time point. This indicates a 12.4-fold increase in 16p fusion events and a 3.6-fold increase in 21q events compared with the parental cell line. Likewise, the maximal frequencies of fusion events for Clone 3 (C3) were  $7.00 \times 10^{-5}$  p/dg and  $1.10 \times 10^{-4}$  p/dg detected with the 16p and 21q probes, respectively. This indicates a 2.7-fold increase for 16p events and a 4.2-fold increase for 21q events compared with the parental cell line.

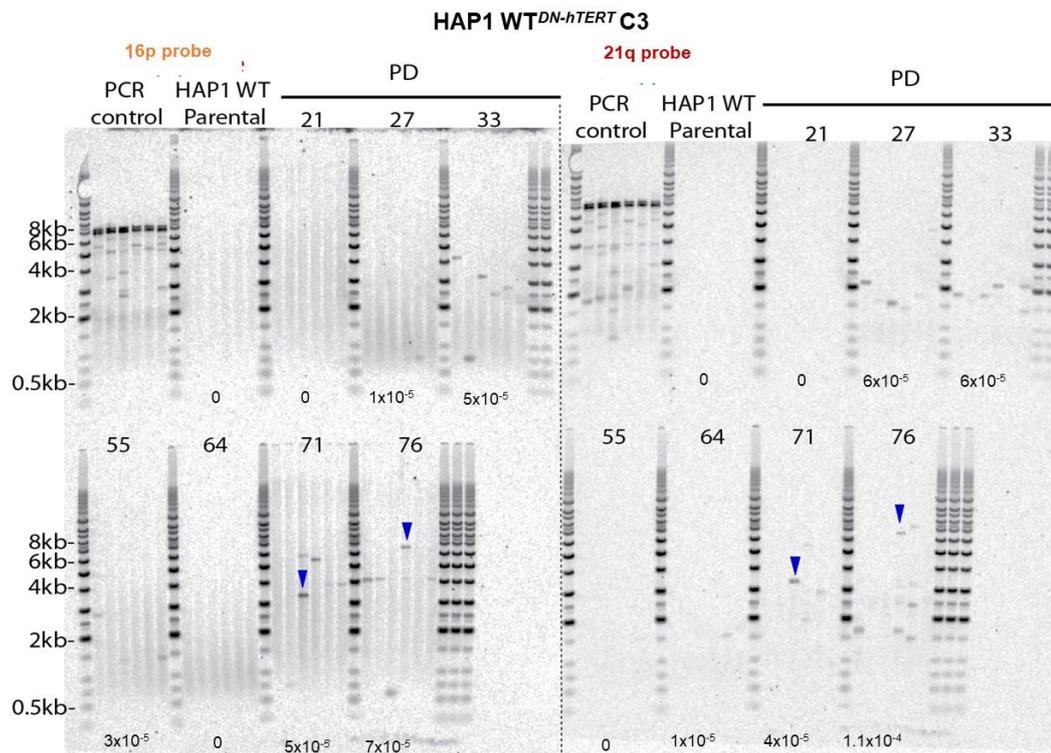
Across the 7 time points sampled for these two clones (between PD 20 – 76), 39 (C2) and 45 (C3) fusion events were detected in total that represent independent *de novo* telomere fusion events using 16p and 21q probes. Events detected with the 16p probe or the 21q probe only are referred to as ‘16p’ or ‘21q’ events whereas events detected with both probes most likely reflect 16p-21q inter-chromosomal events. (**Figure 4.3A-B**). C2 and C3 had comparable numbers of ‘16p’ events (21 and 19 fusion events detected respectively), however, C3 had 40% more ‘21q’ events than C2 (16 and 24 fusion events detected, respectively). Events observed with a single probe may reflect sister chromatid/intra-chromosomal events, although this cannot be confirmed from this assay as it is not possible to determine the particular 16p or 21q family chromosomes involved in each fusion event. Only 2 events were conjointly detected (based on common MW) with the 16p and 21q probes for each clone (5.1% C2 total bands; 4.4% C3 total bands). These events are likely 16p-21q inter-chromosomal

fusion events. Therefore, greater proportions of '16p' and '21q' events were observed compared with 16p:21q inter-chromosomal fusion events for these HAP1 WT<sup>DN-hTERT</sup> clones ( $p \leq 0.001$ , Fisher's exact test) (**Figure 4.3C**). There were no significant differences between the proportions of '16p' or '21q' fusion events in either clone (C2  $p=0.3645$ , C3  $p=0.3988$ , Fisher's exact test). Likewise, there were no significant difference in the proportion of any event type ('16p', '21q' or 16p-21q) when comparing the clones ( $p=0.3814$ ,  $p=0.2818$  and  $p=1.000$ , Fisher's exact test).

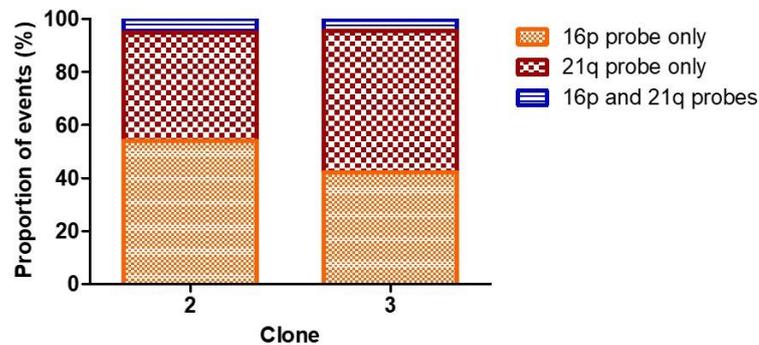
A



B



C



**Figure 4.3:** (A) Fusion profile obtained using the 16p probe (left) and 21q probe (right) for HAP1 WT<sup>DN-hTERT</sup> C2 and C3 (B) following 16p:21q dual-primer PCR. Only 2 identical fusion events are observed on both blots (blue arrows). Fusion frequencies per diploid genome are labelled beneath each sample. (C) Stacked bar chart showing proportions of fusion events detected with the 16p probe only, 21q probe only and events detected with both probes for HAP1 WT<sup>DN-hTERT</sup> C2 and C3 following 16p:21q dual-primer PCR.

#### 4.4.3 Telomere fusion events in HAP1<sup>DN-hTERT</sup> crisis clones

The presence of fusion events was detected with 16p:21q dual-primer PCR for both HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> clones (**Figure 4.4A, one WT and one POLQ clone shown as examples**). The mean maximal fusion frequency per diploid genome is stated based on the equation previously described (*Figure 4.2A*).

##### Fusion frequencies

For the HAP1 WT<sup>DN-hTERT</sup> clones, an 11.9-fold increase in 16p fusion frequency ( $3.58 \times 10^{-5}$  p/dg) and a 15.8-fold increase in 21q fusion frequency ( $1.58 \times 10^{-4}$  p/dg) was detected at the time point of maximal fusions compared with the parental line (*Section 4.4.1*). These values are calculated from only those clones that had crisis-associated telomere fusions (n=4).

An increase in frequency of 16p and 21q fusion events was also detected in the HAP1 POLQ<sup>:DN-hTERT</sup> clones compared with the parental line. A 17.8-fold increase in 16p fusion events ( $1.78 \times 10^{-4}$  p/dg) and a 9.7-fold increase in 21q fusion events ( $1.94 \times 10^{-4}$  p/dg) was observed at the time point of maximal fusion detection specific for each clone compared with the parental line. As with the HAP1 WT<sup>DN-hTERT</sup> cell line, these values are calculated from only those clones that had crisis-associated telomere fusions (n=4).

There was no significant difference in the total abundance of fusion events identified between the HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> cell lines ( $p=0.9768$ , Mann Whitney test) (**Figure 4.4B-C**).

##### Proportions of events

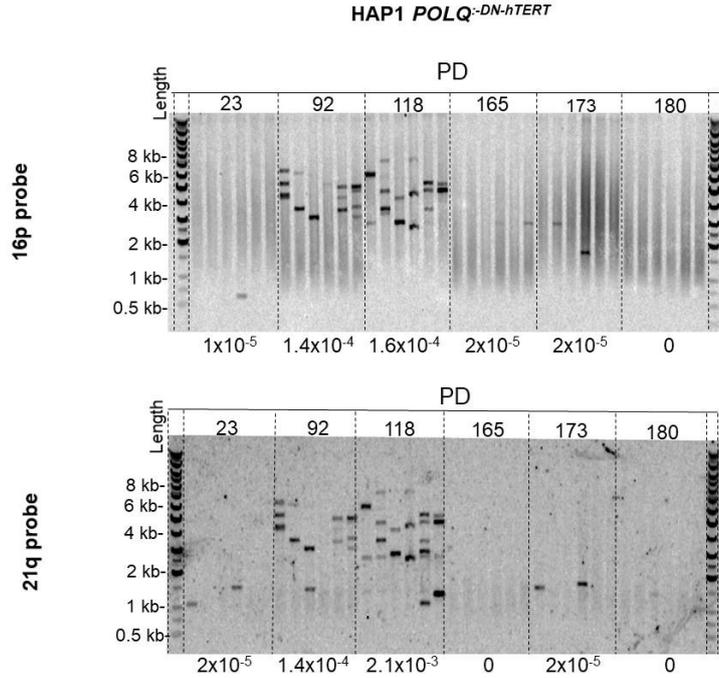
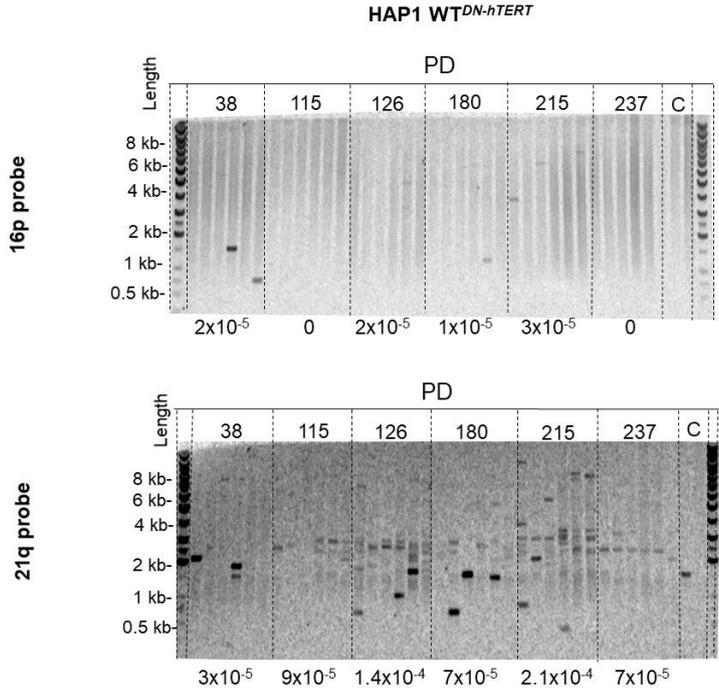
For the HAP1 WT<sup>DN-hTERT</sup> clones, '21q' events ( $1.78 \times 10^{-4}$  p/dg) made up the greatest proportion of fusion events observed (80.8%). A significant 5.4-fold greater abundance ( $p \leq 0.05$ , Mann Whitney test) and a significantly greater proportion

( $p \geq 0.001$ , Fisher's exact test) of '21q' fusion events was recorded for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line ( $3.30 \times 10^{-5}$  p/dg, 13.3%). In contrast, most (84.3%) fusion events observed for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line was 16p-21q inter-chromosomal ( $2.10 \times 10^{-4}$  p/dg). A significant 7.5-fold greater abundance ( $p < 0.05$ , Mann Whitney test) and a significantly greater proportion ( $p \geq 0.001$ , Fisher's exact test) of 16p-21q inter-chromosomal fusion events was observed for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line compared with the HAP1 WT<sup>DN-hTERT</sup> cell line ( $2.79 \times 10^{-5}$  p/dg, 12.7%).

No significant differences were observed when comparing the abundance ( $p = 0.8778$ , Mann Whitney test) or proportions ( $p = 0.1909$ , Fisher's exact test) of '16p' fusion events for the HAP1 WT<sup>DN-hTERT</sup> cell line ( $1.44 \times 10^{-5}$  p/dg, 6.5%) with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line ( $6.00 \times 10^{-6}$  p/dg, 2.4%) (**Figure 4.4D-E**).

Fusion amplicons from HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> clones were sequenced using Illumina paired-end sequencing. This allowed for sequence-analysis of individual fusion events in order to both classify the nature of the fusion events ('16p', '21q' or 16p-21q) and characterise the fusion junctions to try and identify the DNA repair mechanisms involved (*Chapter 5*).

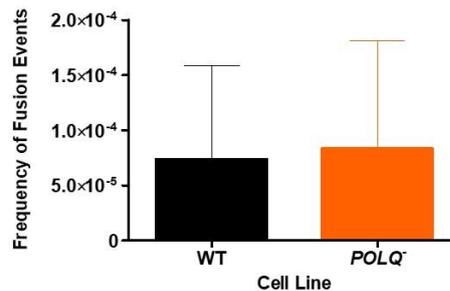
A



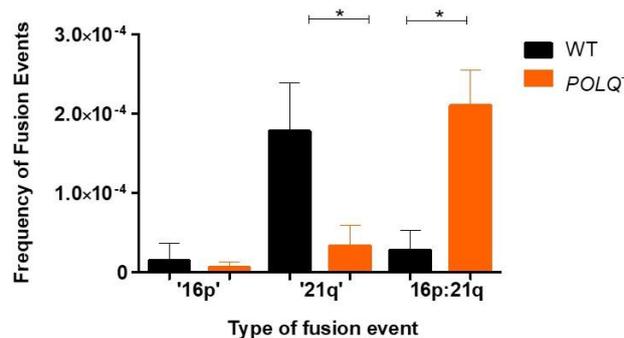
B

Clone	Maximal frequency of fusion events detected per diploid genome		Mean fusion frequency per diploid genome	
	16p	21q	16p	21q
WT 7	$3.00 \times 10^{-5}$	$2.10 \times 10^{-4}$	$3.58 \times 10^{-5}$	$1.58 \times 10^{-4}$
WT 8	$1.80 \times 10^{-5}$	$8.20 \times 10^{-5}$		
WT 9	$5.00 \times 10^{-5}$	$1.70 \times 10^{-4}$		
WT 12	$4.50 \times 10^{-5}$	$1.70 \times 10^{-4}$		
<i>POLQ</i> -3	$1.90 \times 10^{-4}$	$1.90 \times 10^{-4}$	$1.78 \times 10^{-4}$	$1.94 \times 10^{-4}$
<i>POLQ</i> -5	$2.20 \times 10^{-4}$	$2.70 \times 10^{-4}$		
<i>POLQ</i> -8	$1.50 \times 10^{-4}$	$1.75 \times 10^{-4}$		
<i>POLQ</i> -10	$1.50 \times 10^{-4}$	$1.40 \times 10^{-4}$		

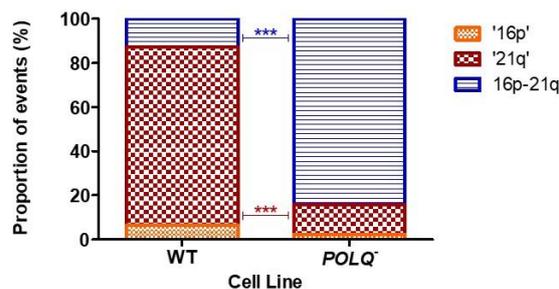
C



D



E



**Figure 4.4:** (A) Fusion profile for HAP1 WT<sup>DN-hTERT</sup> C11 (top) and HAP1 *POLQ*:<sup>DN-hTERT</sup> C8 (bottom) alongside a PCR control (C) following 16p:21q dual-primer PCR detected using 16p and 21q probes. Estimated fusion frequency per diploid genome is shown below each sample (B) Table showing maximal frequency of fusion events per diploid genome detected following 16p:21q PCR and detection using 16p and 21q probes for HAP1<sup>DN-hTERT</sup> cell lines. (C) Bar chart showing total maximal fusion frequencies for 16p:21q fusion events detected with both the 16p and 21q probe for HAP1<sup>DN-hTERT</sup> cell lines. Error bars indicate standard deviation values. (D) Bar chart showing maximal frequency of '16p', '21q' or 16p-21q inter-chromosomal fusion events for HAP1<sup>DN-hTERT</sup> cell lines. Error bars show standard deviation values and significance was determined using a Mann Whitney test (E) Stacked bar chart showing the proportion of '16p', '21q' and 16p-21q fusion events detected for HAP1<sup>DN-hTERT</sup> cell lines. Significance was determined using a Fisher's exact test.

#### 4.4.4 Telomere fusion events in HCT116<sup>DN-hTERT</sup> clones

To determine whether the disruption of *POLQ* caused a change in the abundance of intra- versus inter-chromosomal fusion events in the HCT116 cell line, 17p and 21q fusion primers and radiolabelled probes were used to detect fusion events that arose in HCT116<sup>DN-hTERT</sup> clones passaging through a telomere-driven crisis.

As the 17p subtelomeric sequence is unique, PCR using the 17p primer only amplifies 17p:17p intra-chromosomal fusion events. Therefore, the presence of inter-chromosomal telomere fusion events was detected using 17p and 21q primers, whereas single 17p primer PCR was performed to detect 17p:17p (sister chromatid) intra-chromosomal fusion events. Fusion events with the same molecular weight that were detected by Southern hybridisation using both 17p and 21q probes are considered 17p-21q inter-chromosomal fusion events whereas fusion events unique to the 21q or 17p probe are considered '21q' or 17p:17p events.

#### Frequency of 17p:17p and 17p:21q fusion events in HCT116 parental lines

The equation referenced in (Section 4.4.1) was, again, employed to estimate the frequency of fusion events per diploid genome (Figure 4.2A). For the HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> parental cell lines, no fusion events were detected with the 17p probe following 17p single-primer and 17p:21q dual-primer PCR. However, a mean fusion frequency of  $1.00 \times 10^{-5}$  p/dg was calculated for both HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> parental cell lines when 17p:21q dual-primer amplicons were detected with the 21q probe (Figure 4.5A).

#### Frequency of 17p:17p fusion events in HCT116<sup>DN-hTERT</sup> clones

For the HCT116 WT<sup>DN-hTERT</sup> clones, 17p:17p sister-chromatid events were observed following 17p single-primer PCR ( $1.20 \times 10^{-4}$  p/dg). In contrast, no 17p:17p intra-chromosomal fusion events were detected for the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones at several time points tested (Figure 4.5B) (*The presence of amplifiable DNA in these*

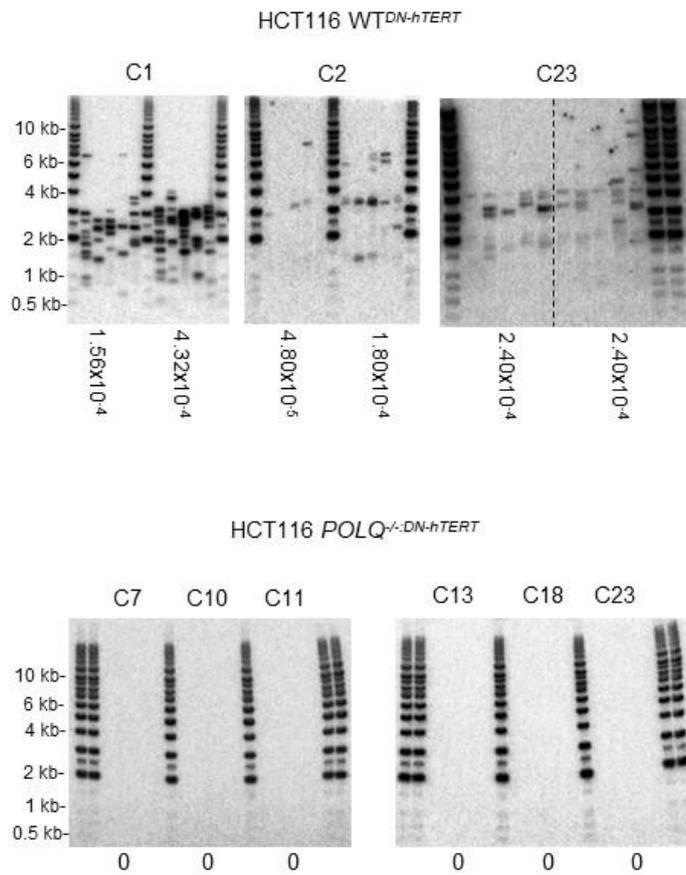
*samples confirmed in Figure 4.5C*). To further investigate the lack of 17p:17p telomere fusion events in these clones, the fusion PCR was repeated using 25, 50 and 100 ng of DNA to confirm that DNA quantity was not affecting the amplification of these products, that is, that no sensitivity threshold was missed. This was also performed using 17p:21q dual-primer PCR to confirm the presence of amplifiable input DNA and the presence of 17p-21q fusion events in these samples. Increasing the input DNA caused a decrease in the abundance of 17p:21q fusion events detected relative to the input DNA. However, no 17p:17p fusion events were detected at 25, 50 or 100 ng. This suggests that the absence of POLQ has an active impact on the generation of 17p:17p intra-chromosomal fusion events. **(Figure 4.5C)**

A

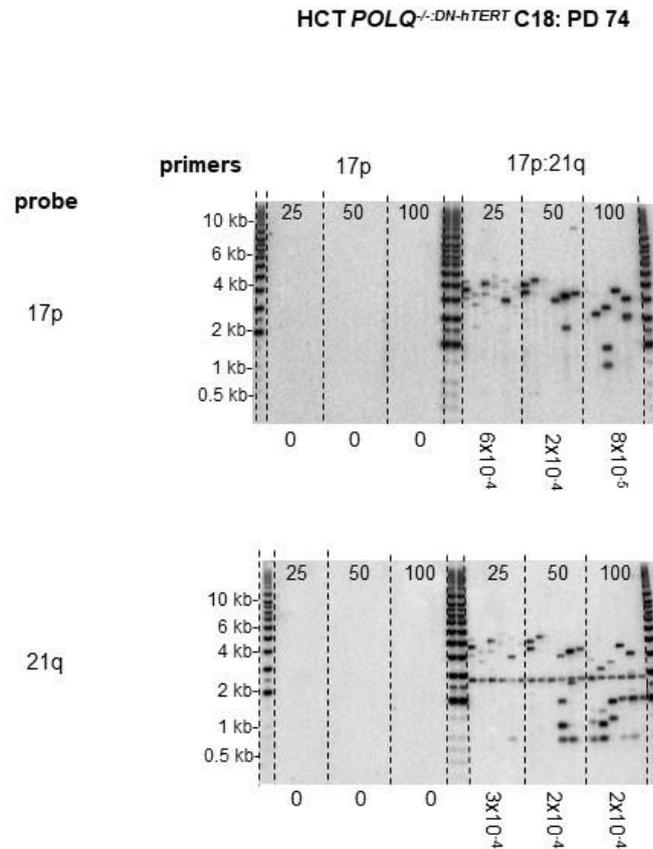
Primers	Probe	Mean fusion frequency per diploid genome			
		Parental		HCT116 <sup>DN-hTERT</sup>	
		WT	<i>POLQ</i> <sup>-/-</sup>	WT	<i>POLQ</i> <sup>-/-</sup>
17p	17p	0	0	1.20x10 <sup>-4</sup>	0
17p:21q	17p	0	0	6.30x10 <sup>-5</sup>	1.31x10 <sup>-4</sup>
	21q	1.00x10 <sup>-5</sup>	1.00x10 <sup>-5</sup>	1.50x10 <sup>-4</sup>	1.74x10 <sup>-4</sup>

B

17p:17p



C



**Figure 4.5:** (A) Table showing 17p:17p and 17p:21q mean fusion frequencies per diploid genome for HCT116 parental and HCT116<sup>DN-hTERT</sup> clones. (B) Fusion profiles for 17p:17p fusion events for HCT116 WT<sup>DN-hTERT</sup> (top) and HCT116 *POLQ*<sup>-/-</sup>*DN-hTERT* (bottom) clones. Fusion frequencies per diploid genome shown beneath each sample (C) 17p:17p (left) and 17p:21q (right) fusion profile for HCT116 *POLQ*<sup>-/-</sup>*DN-hTERT* C18 using 25, 50 and 100 ng input DNA labelled with 17p (top) and 21q probe (bottom). Estimated fusion frequencies per diploid genome shown beneath each sample.

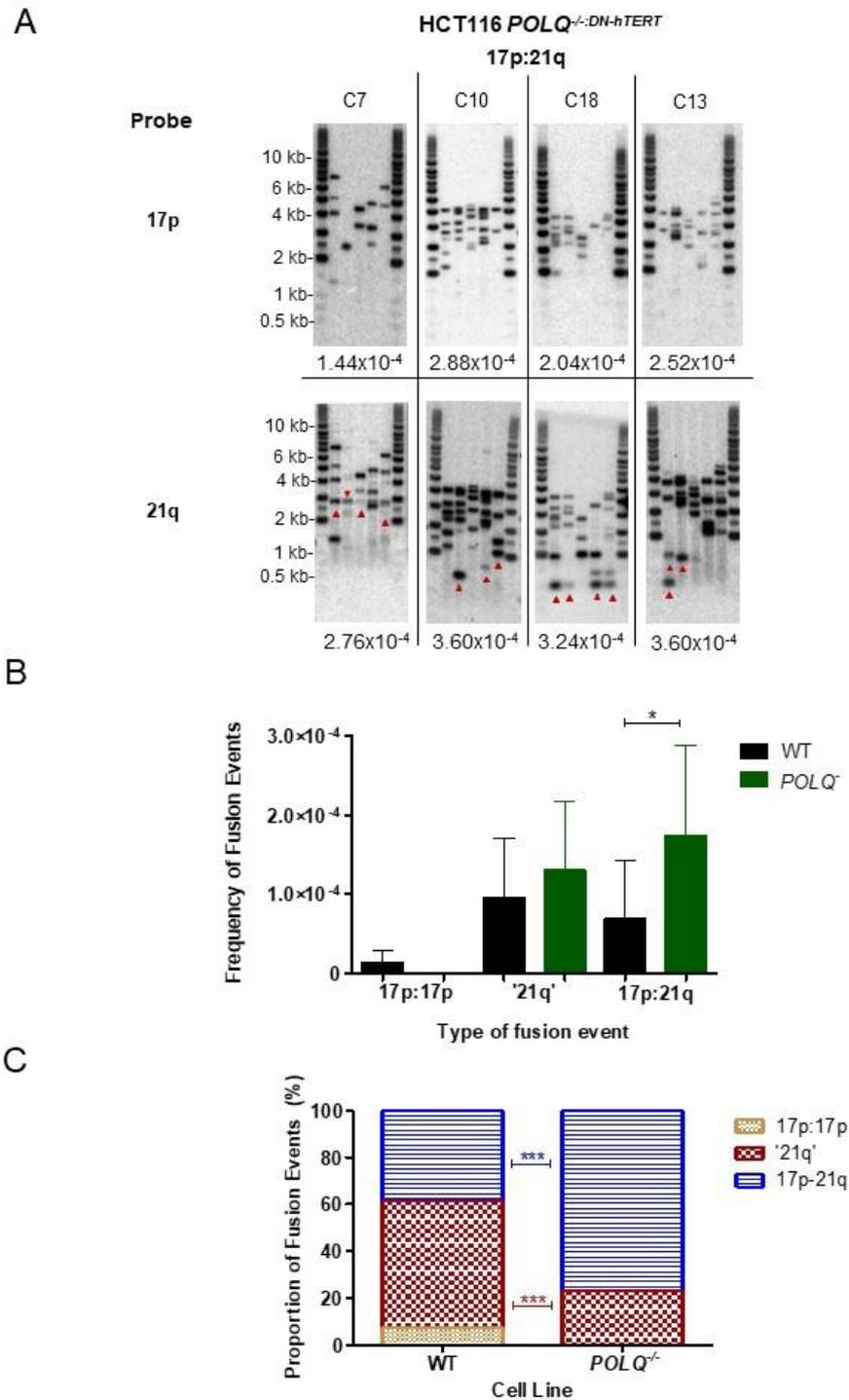
### Comparison of 17p:21q inter-chromosomal fusions in HCT116 WT<sup>DN-hTERT</sup> and *POLQ*<sup>-/-:DN-hTERT</sup> clones

Dual-primer (17p:21q) PCR was performed on HCT116 WT<sup>DN-hTERT</sup> clones and HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones at 5 different sampling time points to detect telomere fusion events associated with telomere-driven crisis (WT n=8, *POLQ*<sup>-/-</sup> n=14) (**Figure 4.6A**). As observed with the HAP1<sup>DN-hTERT</sup> clones, a difference in the type of fusion events was also detected when comparing HCT116 WT<sup>DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones.

For the HCT116 WT<sup>DN-hTERT</sup> clones, the greatest proportion (53.8%) of fusion events observed comprised '21q' events ( $9.60 \times 10^{-5}$  p/dg). Whilst there was no significant difference when comparing the frequency of '21q' fusion events ( $p=0.3271$ , Mann Whitney test) between the clones, a significantly greater proportion ( $p \leq 0.001$ , Fisher's exact test) of '21q' events was detected for the HCT116 WT<sup>DN-hTERT</sup> clones compared with the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones ( $4.20 \times 10^{-5}$  p/dg, 20%) (**Figure 4.6B-C**).

The greatest proportion of fusion events detected in the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones was 17p:21q inter-chromosomal fusion events ( $1.40 \times 10^{-4}$  p/dg, 80%). A significantly greater frequency ( $p < 0.05$ , Mann Whitney test) and proportion ( $p \leq 0.001$ , Fisher's exact test) of 17p-21q inter-chromosomal fusion events was recorded for the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones compared with the HCT116 WT<sup>DN-hTERT</sup> clones ( $6.90 \times 10^{-5}$  p/dg, 38.7%) (**Figure 4.6B-C**).

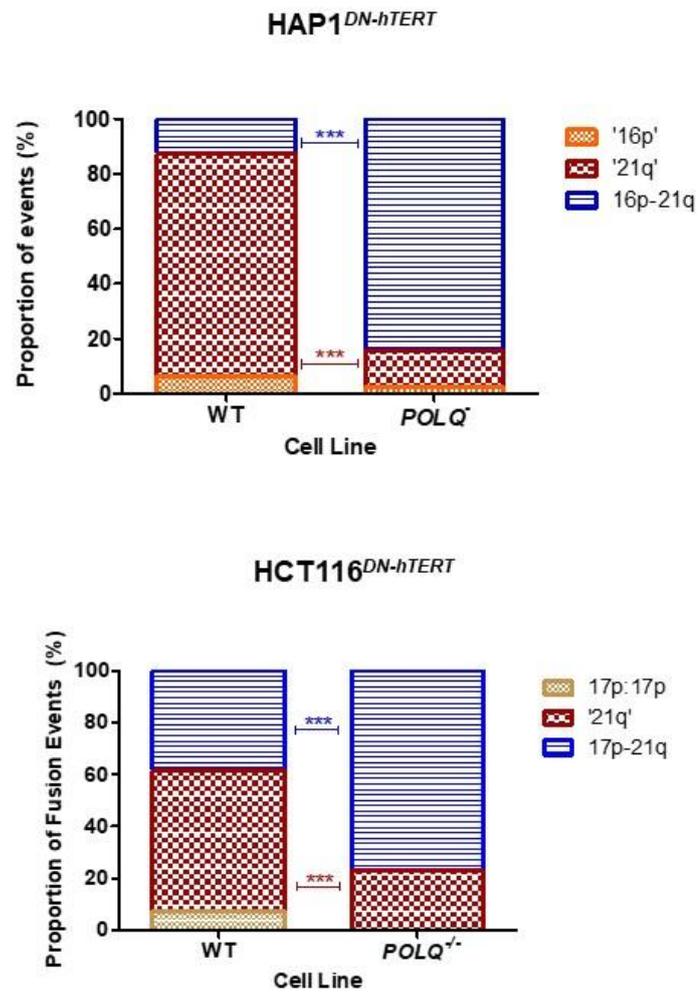
17p:17p fusion events made up 7.6% ( $1.35 \times 10^{-5}$  p/dg) of the total fusion events observed for the HCT116 WT<sup>DN-hTERT</sup> cell line. As no 17p:17p fusion events were detected for the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cell line, a significantly greater proportion of 17p:17p fusion events were detected for the HCT116 WT<sup>DN-hTERT</sup> clones compared with the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones ( $p \leq 0.001$ , Fisher's exact test) (**Figure 4.6B-C**).



**Figure 4.6:** (A) 17p:21q fusion events detected with 17p (top) and 21q probes (bottom) for HCT116 *POLQ*<sup>-/-</sup>:DN-*hTERT* clones. Events unique to 21q probe highlighted by red arrows where possible. Fusion frequencies per diploid genome are shown beneath each sample. (B) Bar chart showing frequency of 17p:17p, '21q' and 17p-21q fusion events for HCT116 WT<sup>DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-</sup>:DN-*hTERT* clones. Significance was determined using a Mann Whitney test. (C) Stacked bar chart showing proportion of 17p:17p, '21q' and 17p-21q events for HCT116 WT<sup>DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-</sup>:DN-*hTERT* clones. Significance was determined using a Fisher's exact test.

#### 4.4.5 Comparing fusion profiles for HAP1<sup>DN-hTERT</sup> and HCT116<sup>DN-hTERT</sup> cell lines passaging telomere-driven crisis

An increase in the proportion of inter-chromosomal fusion events was observed following the disruption of *POLQ* in two distinct cell lines that were passaged through telomere-driven crisis. Whilst it was not possible to discern between inter- and intra-chromosomal fusion events observed using 16p and 21q primers, a reduction in the proportion of '21q' events was observed for the HAP1 *POLQ*<sup>-/-:DN-hTERT</sup> cell line compared with the HAP1 WT<sup>DN-hTERT</sup> cell line. Similarly, no 17p:17p sister-chromatid fusion events were observed for the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cell line whereas these 17p:17p events were detected for the HCT116 WT<sup>DN-hTERT</sup> cell line (**Figure 4.7**). These data therefore suggest that the disruption of *POLQ* is associated with the inhibition of intra-chromosomal fusion events and an increase in inter-chromosomal fusion events.



**Figure 4.7:** Stacked bar charts showing proportion of fusion event types for HAP1<sup>DN-hTERT</sup> (top) and HCT116<sup>DN-hTERT</sup> (bottom) cell lines passing telomere-driven crisis. Increased proportions of inter-chromosomal fusion events (blue) observed for both POLQ-deficient cell lines compared with WT cell lines. Significance was determined using a Fisher's exact test.

#### **4.4.6 Transient Transfection of WT and POLQ-deficient cell lines with subtelomere-targeting TALEN pairs**

As subtelomeric sequences have specialised DNA repair capacity (Muraki et al. 2015), error prone repair processes can be activated following DNA damage resulting in telomere fusion events, even within telomerase-expressing cells with stable telomere lengths. To test this hypothesis, TALEN (Transcription Activator-Like Effector Nuclease) pairs were developed to induce DSBs within subtelomeric regions (Liddiard et al. 2016; Liddiard et al. 2018) and DNA repair was monitored using telomere fusion assays.

In order to further examine the role of POLQ in DNA repair, specific subtelomeric DSBs were induced in HAP1, HCT116 and RPE1<sup>hTERT</sup> parental cell lines by transient nucleofection with TALEN-encoding plasmid pairs. The 17p TALEN pairs were used to induce a DSB in the 17p subtelomere, 14 bp from the start of the telomere. The 16p/21q TALEN pairs were used to induce a DSB 1.5 kb from the start of the telomere repeat arrays of the homologous 16p and 21q telomere families.

For these experiments, cells were also transfected with a GFP plasmid to provide control samples that had been subjected to the same experimental conditions as the cells that underwent nucleofection with TALEN plasmids. In addition, these controls enabled the determination of the efficiency of the transfection (and nucleofection) through detection of GFP expression in viable cells via a Hoechst 33342 and propidium iodide assay (*Chapter 2*). These GFP controls can therefore be used to compare the impact of the TALENs for cell viability and telomere fusion frequency.

#### **4.4.7 Transient Transfection of HAP1 cells with TALEN pairs**

The 17p telomere could not be amplified in either HAP1 WT or HAP1 *POLQ* cells with 17p primers (*Chapter 3*), which could suggest that this cell line has an alternative 17p subtelomeric sequence or location. This also renders the 17p TALEN (that also targets the 17p subtelomere) ineffective in this cell line. Therefore, 16p/21q TALEN

pairs were utilised to induce DSBs in the 16p and 21q family of telomeres. The mean telomere length of the HAP1 WT parental line was 13.5 kb compared with 21.9 kb for the HAP1 *POLQ*<sup>-</sup> parental line (determined by TRF, *Chapter 3*).

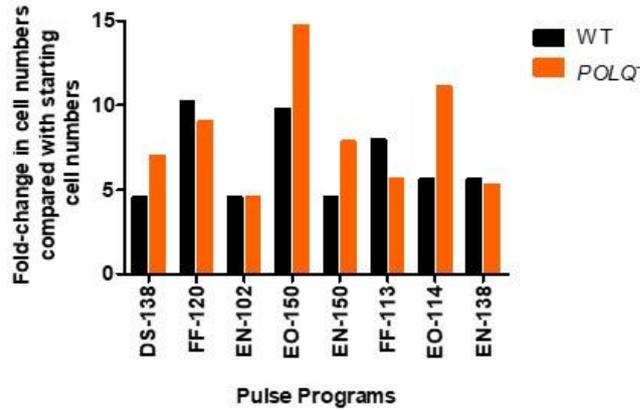
### **Nucleofection Optimisation in HAP1 cells**

To determine the optimum conditions for effective nucleofection of these cells, 8 different nucleofection pulse programs (manufacturer's settings) were tested based on recommended programs for similar cell lines. For each pulse program tested,  $5 \times 10^5$  cells were used. The pulse program, FF-113, was chosen as the most appropriate as it offered the highest transfection ratios (based on GFP expression) without too much of a detrimental effect on cell numbers and viability following transfection (**Figure 4.8A-B**). The mean GFP transfection efficiency for the HAP1 WT cell line across three replica experiments was 72.6% compared to a mean GFP transfection efficiency of 67.3% for the HAP1 *POLQ*<sup>-</sup> cell line (**Figure 4.8C-D**). This similarity in transfection efficiency allows fair comparisons between fusion frequencies observed for the two cell lines as these differences are due to the repair mechanisms utilised by each cell line rather than any difference in the efficiency of the nucleofection experiment. In addition, as the level of protein translation is similar (based on GFP), it is expected that similar levels of functional TALENs will be active in each cell line.

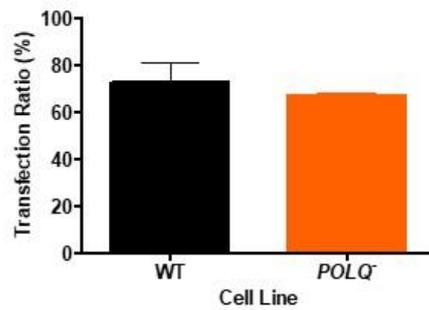
A

Cell Type	Pulse Program															
	DS-138		FF-120		EN-102		E0-150		EN-150		FF-113		E0-114		EN-138	
	V (%)	TR (%)	V (%)	TR (%)	V (%)	TR (%)	V (%)	TR (%)	V (%)	TR (%)	V (%)	TR (%)	V (%)	TR (%)	V (%)	TR (%)
HAP1 WT	89	29	81	50	94	28	73	48	90	57	90	61	86	54	79	28
HAP1 <i>POLQ</i> <sup>-/-</sup>	95	18	86	49	95	16	80	53	86	28	91	48	78	34	90	55

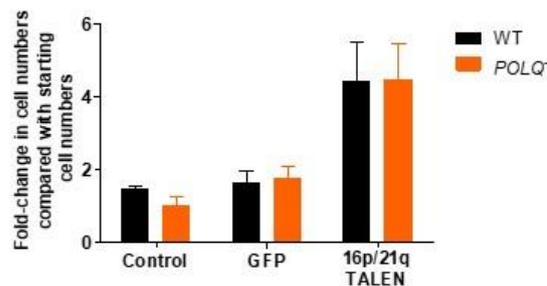
B



C



D



**Figure 4.8:** (A) Table showing viability (V) and transfection ratio (TR) for HAP1 WT and *POLQ*<sup>-/-</sup> cell lines 24 h after nucleofection with GFP plasmids using different pulse programs. (B) Bar chart showing fold-change in cell numbers for HAP1 WT and HAP1 *POLQ*<sup>-/-</sup> cell lines 24 h after nucleofection with GFP plasmids using different pulse programs compared with starting cell numbers. (C) Bar chart showing mean transfection ratios for HAP1 WT and HAP1 *POLQ*<sup>-/-</sup> cell lines 24 h after nucleofection with GFP plasmid. Data from three replica experiments. (D) Bar chart showing fold-change in cell numbers for HAP1 WT and HAP1 *POLQ*<sup>-/-</sup> cell lines 24 h after nucleofection with GFP plasmids or 21q TALENs using the FF-113 pulse program, in addition to untransfected controls compared with starting cell numbers. Data from three replica experiments.

## Abundance of Fusion Events

### Fusion events in parental cell lines

To analyse any differences in DNA repair activity 24 h following nucleofection with 16p/21q TALEN pairs, 16p and 21q primers were used separately and in combination to detect telomere fusion events. The equation described previously (*Figure 4.2A*) was also used to estimate the frequency of fusion events per haploid (HAP1 *POLQ*<sup>-</sup> parental) or diploid (HAP1 WT parental) genome for the parental cell lines.

In order to set the baseline for comparison with TALEN activity, fusion analysis was performed for untransfected HAP1 WT and HAP1 *POLQ*<sup>-</sup> parental cell lines. No fusion events were detected in either cell line following 16p and 21q single-primer fusion PCR or 16p:21q dual-primer PCR.

### 16p and 21q single-primer PCR

An increase in fusion events was detected for both HAP1 WT and HAP1 *POLQ*<sup>-</sup> cell lines that had undergone nucleofection with the 16p/21q TALEN pairs when compared to GFP and untransfected controls. This increase was observed following fusion PCR with 16p and 21q primers separately and together.

No '16p' or '21q' fusion events were observed for either the untransfected or GFP controls. However, the mean '16p' fusion frequency was  $6.76 \times 10^{-6}$  per haploid/diploid genome for both the HAP1 WT and HAP1 *POLQ*<sup>-</sup> cell lines transfected with the 16p/21q TALEN pairs. In addition, a greater number of '21q' fusion events was detected for the HAP1 WT cells ( $8.67 \times 10^{-5}$  p/dg) compared with HAP1 *POLQ*<sup>-</sup> cells ( $1.33 \times 10^{-5}$  p/hg) (**Figure 4.9A**).

### 16p:21q dual-primer PCR

A greater abundance of fusion events was detected in both cell lines following dual-primer 16p:21q PCR compared with single-primer PCR.

Most fusion events detected were inter-chromosomal telomere fusion events for both the HAP1 WT ( $2.07 \times 10^{-4}$  p/dg, 78%) and HAP1 *POLQ*<sup>-</sup> ( $3.07 \times 10^{-4}$  p/hg, 80%) cell line. These events were detected with both probes. There was no significant difference in the abundance ( $p=0.2000$ , Mann Whitney test) or proportion ( $p=0.6951$ , Fisher's exact test) of 16p-21q inter-chromosomal fusion events recorded for the two cell lines.

Similarly, there was no significant difference in the abundance ( $p=0.8248$ , Mann Whitney test) or proportion ( $p=1.00$ , Fisher's exact test) of '21q' fusion events when comparing the HAP1 WT cell line ( $6.00 \times 10^{-5}$  p/dg, 17%) with the HAP1 *POLQ*<sup>-</sup> cell line ( $5.33 \times 10^{-5}$  p/hg, 15%).

Likewise, there was no significant difference in the abundance ( $p=1.00$ , Mann Whitney test) or proportion ( $p=1.00$ , Fisher's exact test) of '16p' fusion events when comparing the HAP1 WT cell line ( $1.33 \times 10^{-5}$  p/dg, 5%) with the HAP1 *POLQ*<sup>-</sup> cell line ( $2.00 \times 10^{-5}$  p/hg, 5%).

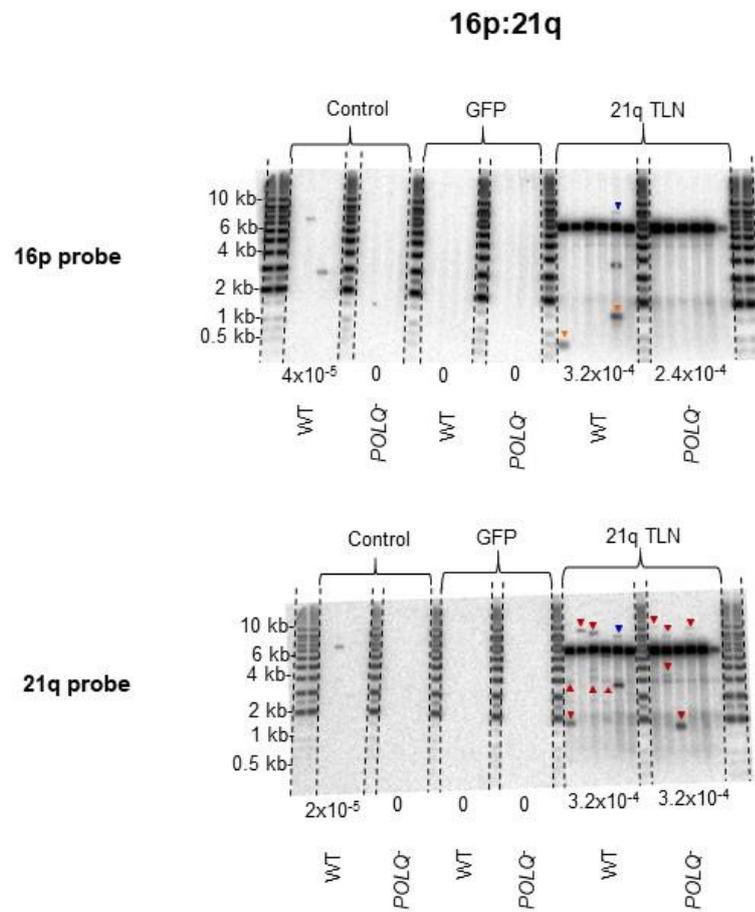
(The untransfected HAP1 WT control had a 16p fusion frequency of  $4 \times 10^{-5}$  p/dg and a 21q fusion frequency of  $2 \times 10^{-5}$  p/dg.)

Overall, 16p-21q inter-chromosomal fusion events were more abundant for both cell lines. These events therefore represent a significantly greater proportion of events than '16p' and '21q' events for both the HAP1 WT cell line ( $p \leq 0.01$ , Fisher's exact tests) and the HAP1 *POLQ*<sup>-</sup> cell line ( $p \leq 0.001$  and  $p \leq 0.01$  respectively, Fisher's exact test) (**Figure 4.9B-D**).

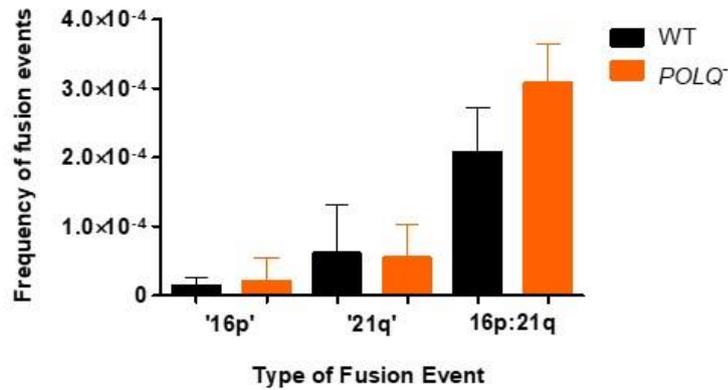
A

Primers	Probe	16p/21q TALEN	
		Mean fusion frequency	
		WT	<i>POLQ</i> <sup>-/-</sup>
16p	16p	$6.67 \times 10^{-6}$	$6.67 \times 10^{-6}$
21q	21q	$8.67 \times 10^{-5}$	$1.33 \times 10^{-5}$
16p:21q	16p	$2.40 \times 10^{-4}$	$2.27 \times 10^{-5}$
	21q	$2.87 \times 10^{-4}$	$3.20 \times 10^{-4}$

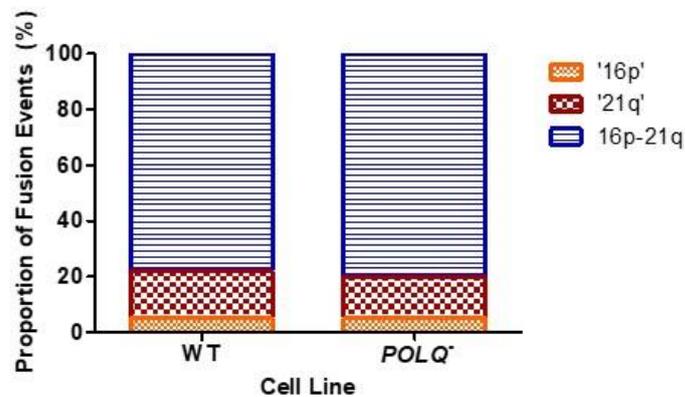
B



C



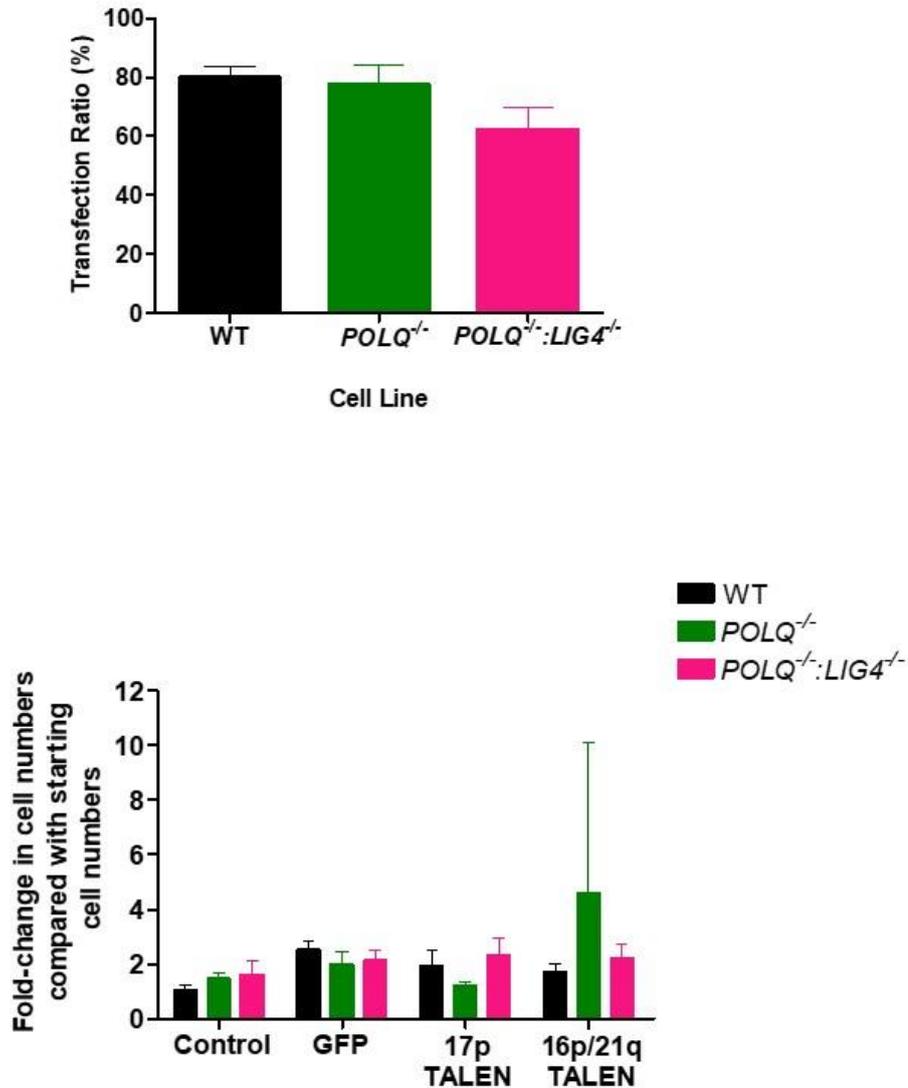
D



**Figure 4.9:**(A) Table showing frequency of fusion events per genome detected with 16p and 21q probes for single-primer and 16p:21q dual-primer PCR for HAP1 WT and HAP1 *POLQ*<sup>-</sup> cells that had undergone nucleofection with 16p:21q TALENs (TLN) (B) Fusion blots hybridised with 16p probe (top) and 21q probe (bottom) following 16p:21q dual-primer PCR for HAP1 WT and HAP1 *POLQ*<sup>-</sup> cell lines following nucleofection with GFP or 16p:21q TALEN plasmids as well as untransfected controls. Fusion frequencies per genome shown beneath each sample. Fusion events identified with 16p probe only highlighted by orange arrows, 21q only fusion events highlighted by red arrows. (C) Chart showing frequency of '16p', '21q' and 16p-21q inter-chromosomal fusion events generated following 16p:21q primer PCR for HAP1 WT and HAP1 *POLQ*<sup>-</sup> cell lines. Error bars show standard deviation and significance was determined using a Mann Whitney test. (D) Stacked bar chart showing the proportion of '16p', '21q' and 16p:21q fusion events following 16p:21q dual-primer PCR for HAP1 WT and HAP1 *POLQ*<sup>-</sup> cell lines that had undergone nucleofection with 16p/21q TALEN pairs.

#### 4.4.8 Transient Transfection of HCT116 cells with TALEN pairs

Both 17p and 16p/21q TALEN pairs were utilised to induce specific subtelomeric DSBs within HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cell lines. As this cell line had previously been interrogated using this system, no optimisation was required, so the DS-138 pulse program was used for the nucleofection of these cell lines. As HAP1 cell growth is more rapid than HCT116 cell lines, a 48 h harvest time point was chosen to analyse DNA repair activity following nucleofection with TALEN pairs for the HCT116 cell line. In addition to the HCT116 WT and *POLQ*<sup>-/-</sup> cell lines, a HCT116 cell line in which both *POLQ* and *LIG4* had been disrupted (Professor Eric Hendrickson, *Chapter 2*) was also used. The disruption of both genes should result in the abrogation of both classical and alternative NHEJ pathways. The mean GFP transfection efficiency for the HCT116 WT cell line was 80% compared to a mean GFP transfection efficiency of 77.6% for the HCT116 *POLQ*<sup>-/-</sup> cell line. However, the HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line had a lower mean GFP transfection efficiency of 62.3% (**Figure 4.10**). The mean telomere length of the HCT116 WT cell line was 5.3 kb compared with mean telomere lengths of 8.0 kb and 5.4 kb for the HCT116 *POLQ*<sup>-/-</sup> and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines (telomere lengths derived from TRF analysis, *Chapter 3*).



**Figure 4.10:** Bar chart showing mean transfection ratios for HCT116 WT, HCT116 *POLQ*<sup>-/-</sup> and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines 48 h following nucleofection with GFP plasmids (top) Data obtained from three replica experiments. Bar chart showing fold-change in cell numbers for HCT116 WT, HCT116 *POLQ*<sup>-/-</sup> and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines 48 h following nucleofection with GFP, 17p or 16p/21q TALEN pairs alongside untransfected controls compared with starting cell numbers (bottom). Data obtained from three replica experiments.

### Transfection using the 17p TALENs

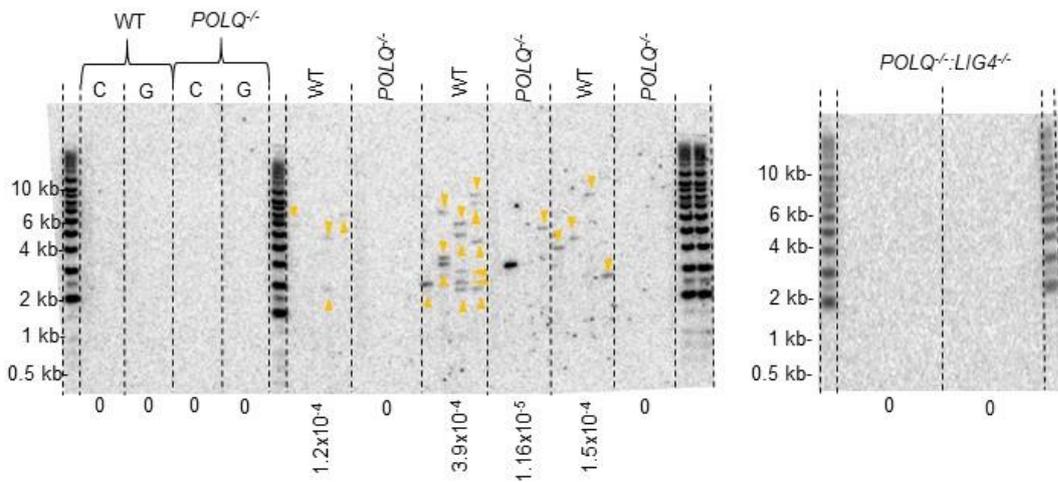
For the HCT116 WT cell line, an increase in the abundance of fusion events involving the 17p telomere was detected 48 h following nucleofection with the 17p TALEN pairs when compared with GFP and untransfected controls. These fusion events were observed following 17p single-primer PCR so likely consist of sister chromatid fusion events. The equation shown in *Figure 4.2A* was also used to estimate the frequency of fusion event per diploid genome.

The mean frequency of 17p:17p sister chromatid fusion events observed for the HCT116 WT cell line was  $2.00 \times 10^{-4}$  p/dg. In contrast, significantly fewer ( $p < 0.05$ , Mann Whitney test) 17p:17p sister chromatid fusion events were detected for the HCT116 *POLQ*<sup>-/-</sup> cell line following nucleofection with 17p TALEN pairs. The mean fusion frequency for this cell line was  $1.75 \times 10^{-5}$  p/dg. For the HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line, no 17p:17p sister chromatid fusion events were detected following transfection with the 17p TALEN pair (**Figure 4.11**). The presence of amplifiable DNA in these samples is confirmed in *Figure 4.12*.

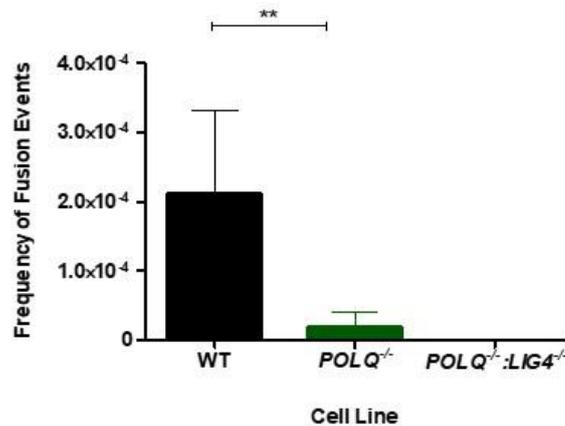
A

		Mean fusion frequency per diploid genome				
Primer	Probe				17p TALEN	
		Control	GFP	WT	<i>POLQ</i> <sup>-/-</sup>	<i>POLQ</i> <sup>-/-</sup> : <i>LIG4</i> <sup>-/-</sup>
17p	17p	0	0	2.00x10 <sup>-4</sup>	1.75x10 <sup>-5</sup>	0

B



C



**Figure 4.11:** (A) Table showing mean fusion frequency of 17p:17p sister chromatid fusion events per diploid genome for HCT116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines 48 h following nucleofection using 17p TALEN pairs alongside GFP and untransfected controls. (B) Fusion blot showing 17p:17p sister chromatid fusion events for HCT116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines 48 h following nucleofection using 17p TALEN pairs in addition to GFP (G) and untransfected controls (C). Fusion frequencies per genome are listed below each sample. (C) Chart showing mean frequency of 17p:17p sister chromatid fusion events per genome for HCT116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines 48h following nucleofection using 17p TALEN pairs. Error bars indicate standard deviation and significance was determined using a Mann Whitney test.

### 17p:21q Fusion Events

Following 17p:21q dual-primer PCR, an increase in fusion events was observed for the HCT116 WT and HCT116 *POLQ*<sup>-/-</sup> cell line samples that had undergone nucleofection with 17p TALEN pairs compared with GFP and untransfected controls. This increase was detected following hybridisation with both the 17p and 21q probes. Events detected with both probes (based on MW) likely represent 17p-21q inter-chromosomal fusion events. The mean frequency of fusion events observed with the 17p probe for the HCT116 WT cell line was  $2.55 \times 10^{-4}$  p/dg compared to  $1.50 \times 10^{-5}$  p/dg for the HCT116 *POLQ*<sup>-/-</sup> cell line. However, as observed with the 17p primer only, no 17p fusion events were detected for the HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line following 17p:21q dual-primer PCR.

In contrast, a greater abundance of 21q events ( $8.00 \times 10^{-5}$  p/dg) was detected for the HCT116 *POLQ*<sup>-/-</sup> cell line compared with the HCT116 WT cell line ( $1.50 \times 10^{-5}$  p/dg) following 17p:21q dual-primer PCR. In addition, the mean frequency of '21q' fusion events detected for the HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line was  $1.00 \times 10^{-5}$  p/dg. Although differences in the abundance of 17p and 21q events was detected for the different cell lines, these differences were not statistically significant (Mann Whitney test). **(Figure 4.12A-C).**

### Proportions of 17p:21q events

Differences in the proportions of 17p:17p, '21q' and 17p-21q inter-chromosomal fusion events for the three cell lines was also detected.

Most fusion events observed following 17p:21q PCR for the HCT116 WT cell line were 17p:17p sister chromatid events (94%,  $2.40 \times 10^{-4}$  p/dg). As such, a significantly greater proportion of 17p:17p events was detected for the HCT116 WT cell line compared to 17p:21q inter-chromosomal fusion events (0%) and '21q' events (6%,  $1.50 \times 10^{-4}$  p/dg) ( $p \leq 0.001$ , Fisher's exact test).

In contrast, the greatest proportion of fusion events observed for the HCT116 *POLQ*<sup>-/-</sup> cell line following nucleofection with 17p TALEN pairs was '21q' fusion events (47%,  $3.50 \times 10^{-5}$  p/dg). However, there was no significant difference in the proportion of '21q' events when compared with both 17p:17p fusion events (20%,  $1.50 \times 10^{-5}$  p/dg) and 17p:21q (33%,  $2.50 \times 10^{-5}$  p/dg) fusion events ( $p = 0.100$ , Fisher's exact test).

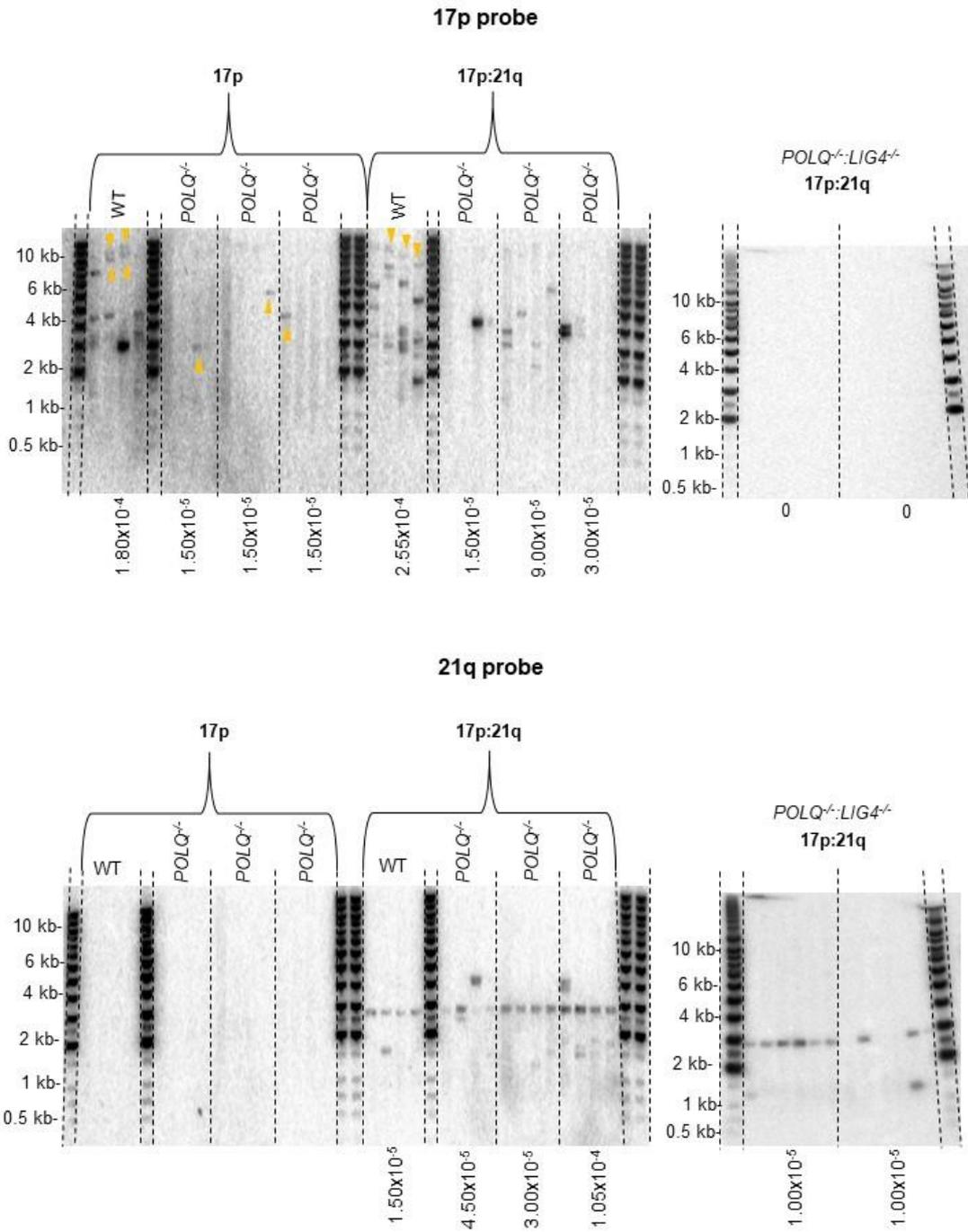
Following nucleofection using 17p TALEN pairs, the only fusion events detected for the HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line were detected with the 21q probe (**Figure 4.12D-E**).

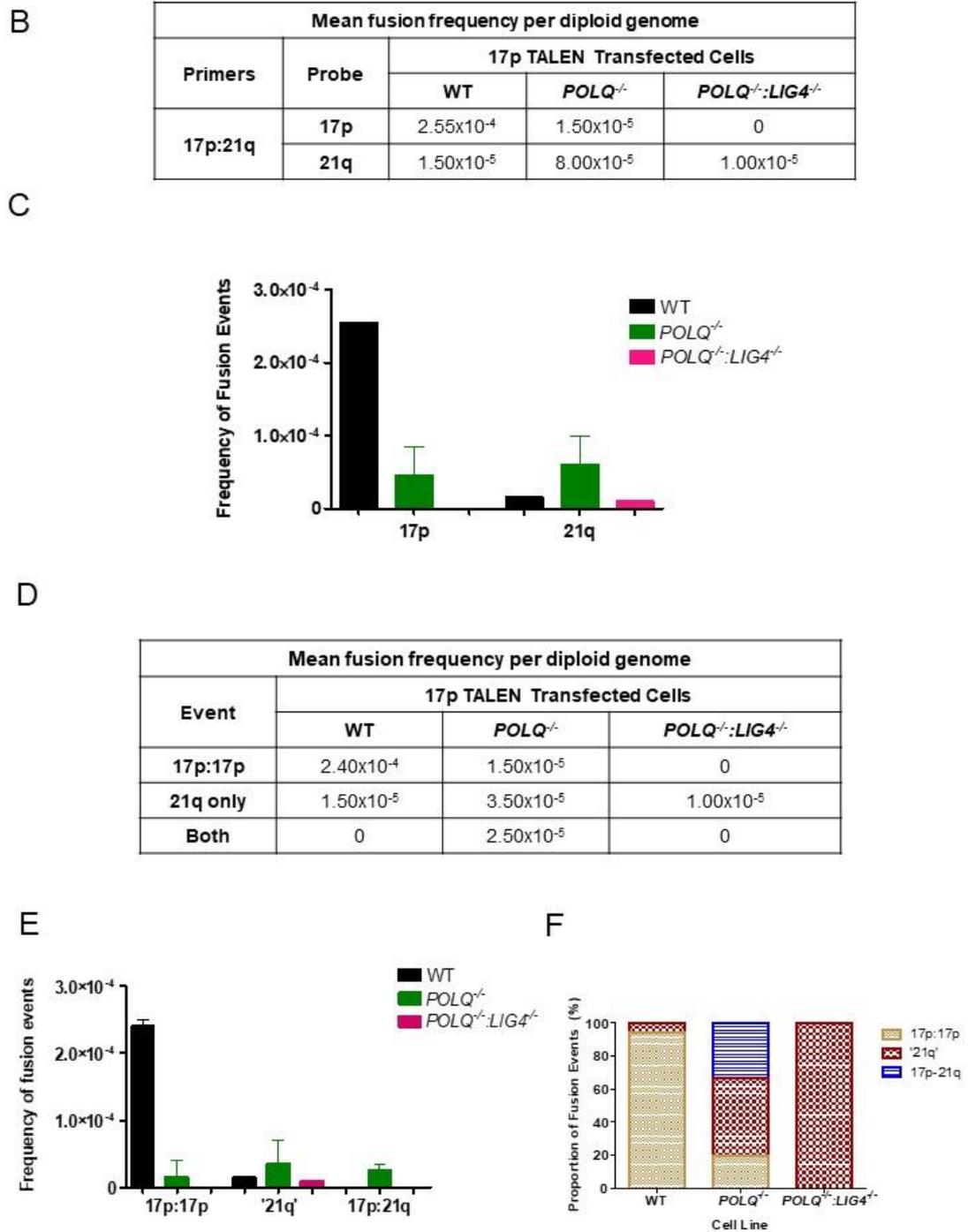
A 16-fold greater abundance of 17p:17p fusion events was observed for HCT116 WT cells which had undergone nucleofection with 17p TALENs when compared with the HCT116 *POLQ*<sup>-/-</sup> cell line. As such, these 17p:17p events represented a significantly greater proportion of events compared to the HCT116 *POLQ*<sup>-/-</sup> and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line ( $p \leq 0.001$ , Fisher's exact test).

In contrast, the greatest proportion of fusion events observed for both HCT116 *POLQ*<sup>-/-</sup> and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines were '21q' events. Therefore, the proportion of '21q' events for both HCT116 *POLQ*<sup>-/-</sup> and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines was significantly greater than the proportion of '21q' events for the HCT116

WT cell line ( $p \leq 0.001$ , Fisher's exact test). Although 17p:21q inter-chromosomal fusion events were only detected for the HCT116 *POLQ*<sup>-/-</sup> cell line, the differences in the proportion of these inter-chromosomal fusion events when compared with the HCT116 WT cell line and the HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line was not statistically significant ( $p=0.0761$ , Fisher's exact tests) (**Figure 4.12F**).

A





**Figure 4.12:** (A) Fusion blots produced following 17p single-primer and 17p:21q dual-primer PCR for HCT 116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cells 48 h following nucleofection with 17p TALEN pairs. Fusion blot hybridised with the 17p (top) and 21q probe (bottom). Fusion frequencies per diploid genome listed beneath each sample. (B) Table and bar chart (C) showing mean fusion frequencies per genome for HCT 116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cells that had undergone nucleofection with 17p TALEN pairs following 17p:21q dual-primer PCR, observed with 17p and 21q probes. Error bars indicate standard deviation values. (D) Table and bar chart (E) showing mean 17p:17p, '21q' and 17p:21q fusion frequencies per genome for HCT 116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cells that had undergone nucleofection with 17p TALEN pairs observed following 17p:21q dual-primer PCR. Error bars indicate standard deviation values. (F) Stacked bar chart showing proportions of 17p:17p, '21q' and 17p:21q fusion events for HCT 116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cells that had undergone nucleofection with 17p TALEN pairs observed following 17p:21q dual-primer PCR.

### Transfection using 16p/21q TALEN pairs

In addition to the 17p TALEN, 16p/21q TALEN pairs were also utilised to induce DSBs within the subtelomeric sequence of these telomere families in the HCT116 lines. To analyse any telomere fusion events induced by these TALEN pairs, 16p and 21q primers were used separately and together to amplify any telomere fusion events involving these telomere families.

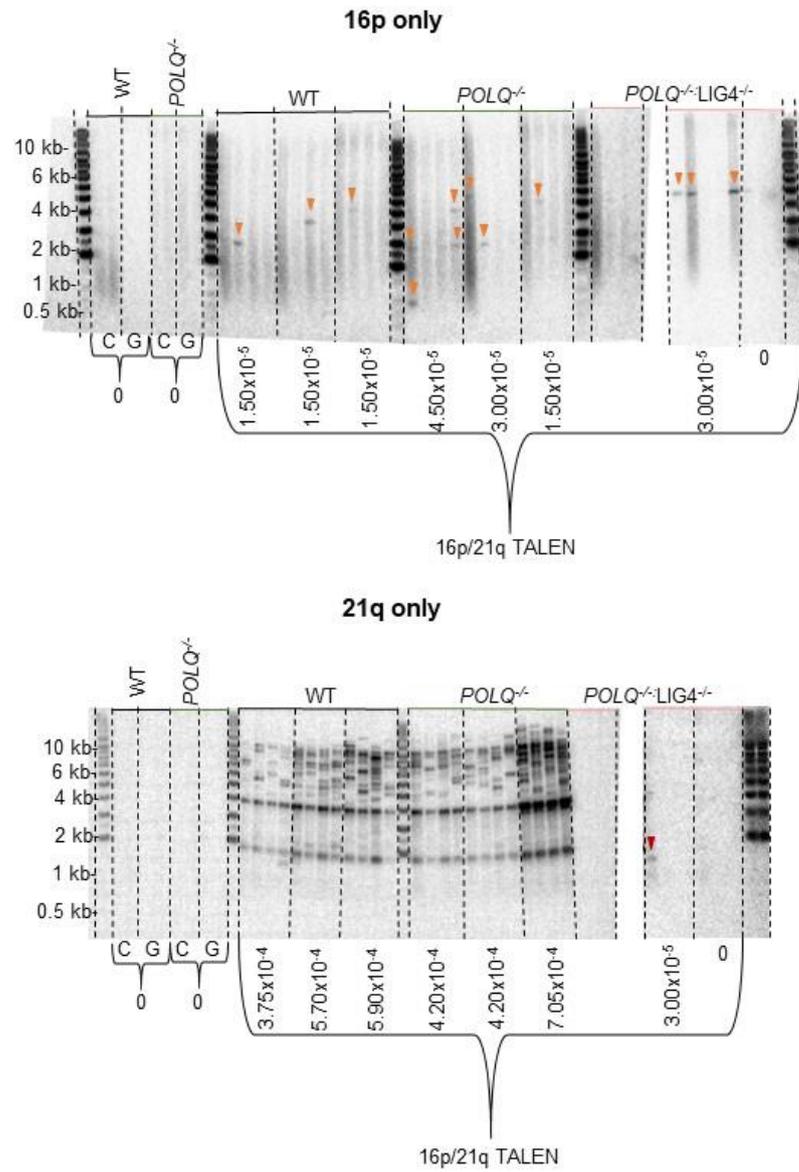
#### 16p single-primer PCR

Following nucleofection with 16p/21q TALEN pairs, an increase in the abundance of telomere fusion events was observed compared with untransfected controls and cells that had been transfected with GFP plasmids. Following 16p single-primer PCR, a greater abundance of fusion events was observed for the HCT116 *POLQ*<sup>-/-</sup> cell line ( $2.50 \times 10^{-5}$  p/dg) compared with both the HCT116 WT ( $1.50 \times 10^{-5}$  p/dg) and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line ( $2.00 \times 10^{-5}$  p/dg) respectively. There were no significant differences in the abundance of '16p' fusion events when comparing the cell lines ( $p=0.100$ , Mann Whitney test) (**Figure 4.13**).

#### 21q single-primer PCR

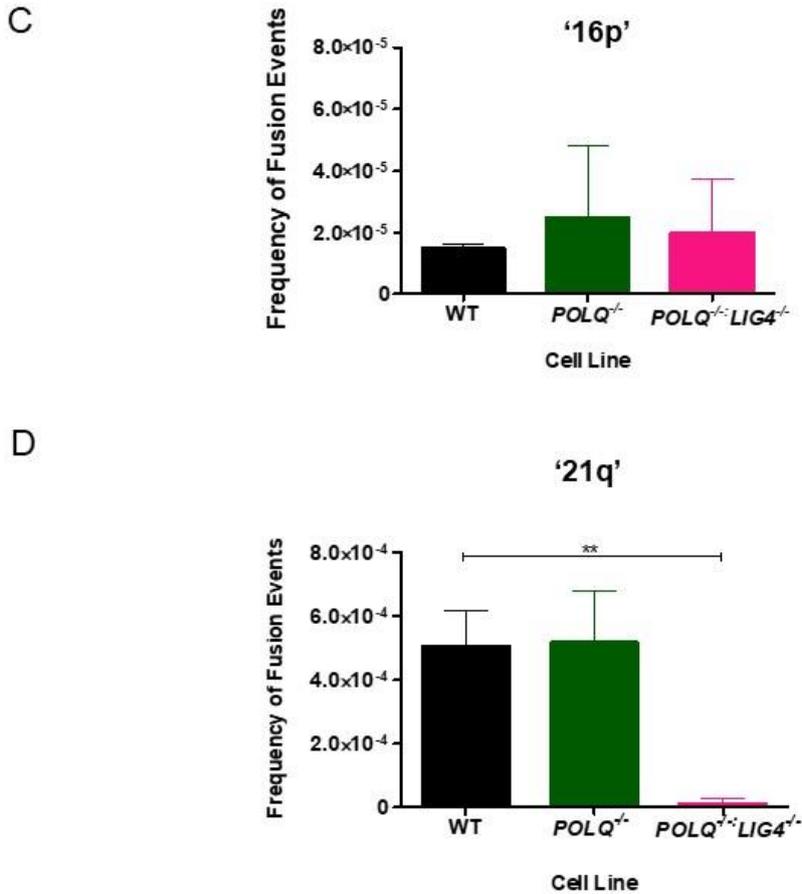
Telomere fusion events were also observed following 21q single-primer PCR. However, as with the '16p' events, there was no significant difference in the frequency of '21q' fusion events for the HCT116 WT cell line compared to the HCT116 *POLQ*<sup>-/-</sup> cell line following nucleofection with 16p/21q TALEN pairs ( $p=0.100$ , Mann Whitney test). In contrast, the mean frequency of '21q' fusion events calculated for the HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line was  $1.00 \times 10^{-5}$  p/dg. Thus, significantly fewer '21q' fusion events were observed for this cell line when compared to the HCT116 WT (5.1-fold) and the *POLQ*<sup>-/-</sup> (5.2-fold) cell lines ( $p \leq 0.01$ , Mann Whitney tests) (**Figure 4.13**).

A



B

Primers	Probe	16p/21q TALEN transfected cells		
		WT	<i>POLQ</i> <sup>-/-</sup>	<i>POLQ</i> <sup>-/-</sup> : <i>LIG4</i> <sup>-/-</sup>
16p	16p	$1.50 \times 10^{-5}$	$2.50 \times 10^{-5}$	$2.00 \times 10^{-5}$
21q	21q	$5.05 \times 10^{-4}$	$5.15 \times 10^{-4}$	$1.00 \times 10^{-5}$



**Figure 4.13: (A)** Fusion blot following 16p primer (top) and 21q primer (bottom) PCR for HCT116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>*LIG4*<sup>-/-</sup> cell lines 48 h following nucleofection with 16p/21q TALEN pairs alongside GFP (G) and untransfected controls (C). Fusion frequencies per genome listed below each sample **(B)** Table showing frequency per genome of '16p' and '21q' fusion events for HCT116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>*LIG4*<sup>-/-</sup> cell lines following nucleofection with 16p/21q TALEN pairs. **(C)** Chart showing '16p' and '21q' **(D)** fusion events for HCT116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>*LIG4*<sup>-/-</sup> cell lines 48 h following nucleofection with 16p/21q TALEN pairs. Error bars indicate standard deviation and significance was determined using a Mann Whitney test.

### 16p:21q fusion events

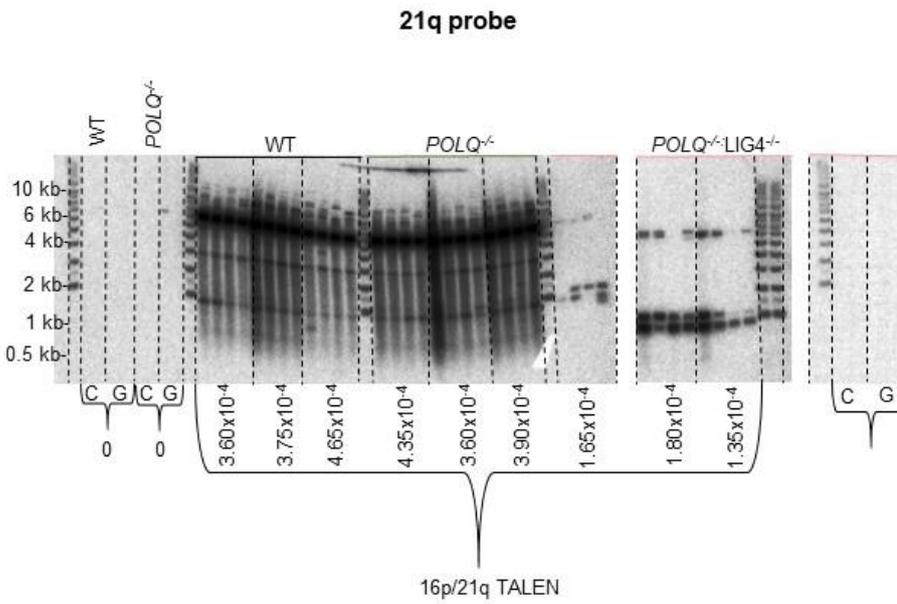
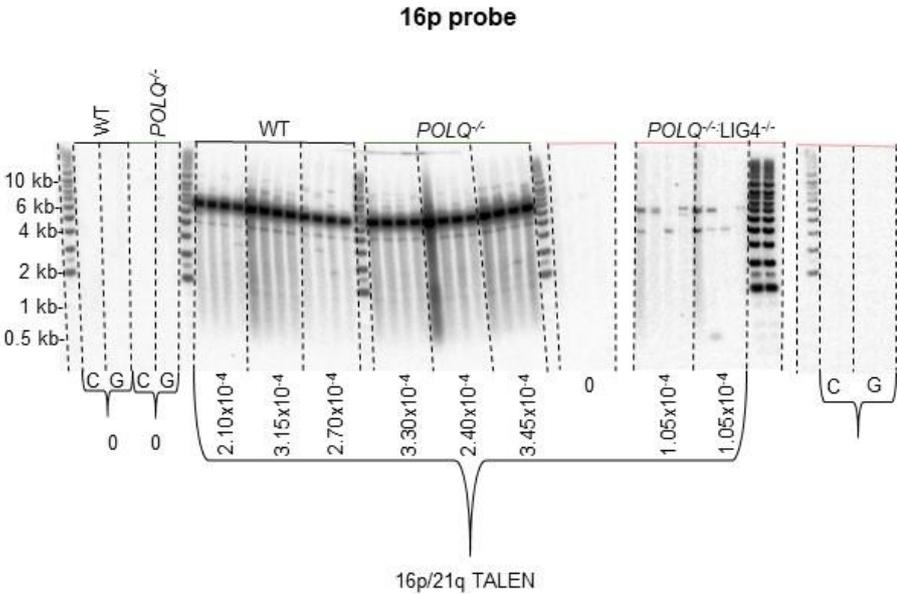
For the HCT116 WT and HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines the greatest proportion of fusion events observed following 16p:21q dual-primer PCR (**Figure 4.14A**) was '21q' events ( $2.75 \times 10^{-4}$  p/dg, 53% and  $5.25 \times 10^{-5}$  p/dg, 54% respectively). For the HCT116 WT cell line there was no significant difference between the proportion of these '21q' events compared with 16p-21q fusion events ( $1.85 \times 10^{-4}$  p/dg, 36%).

Although the greatest proportion of fusion events detected for the HCT116 *POLQ*<sup>-/-</sup> cell line was 16p:21q inter-chromosomal fusion events ( $2.15 \times 10^{-4}$  p/dg, 50%), there was no significant difference between the proportion of these 16p-21q events compared with '21q' events ( $1.55 \times 10^{-4}$ , 36%). However, for the HCT116 *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell line the proportion of 16p:21q inter-chromosomal fusion events ( $5.25 \times 10^{-5}$  p/dg, 25%) was significantly smaller ( $p < 0.05$ , Fisher's exact test) than the proportion of '21q' events.

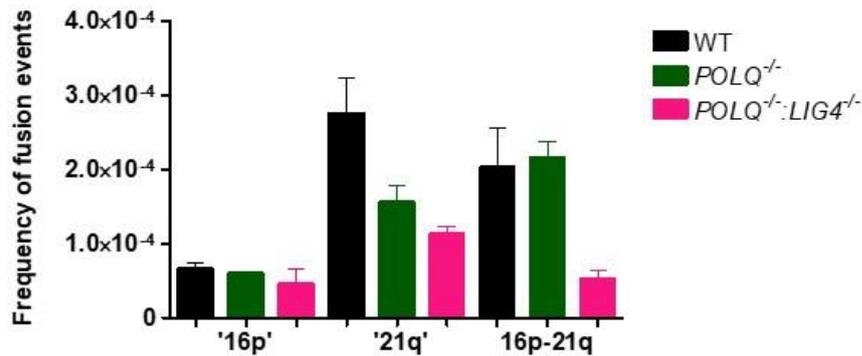
For all three cell lines, the smallest proportion of events observed were '16p' events (WT= $6.50 \times 10^{-5}$  p/dg, 13%, *POLQ*<sup>-/-</sup>= $6.00 \times 10^{-5}$  p/dg, 14% and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup>= $4.50 \times 10^{-5}$  p/dg, 21%). Therefore, these 16p events represented a significantly smaller proportion of events compared with both '21q' events ( $p \leq 0.001$ ,  $p \leq 0.01$  and  $p < 0.05$ , respectively, Fisher's exact test) and 16p:21q events ( $p < 0.05$ ,  $p \leq 0.01$ , respectively, Fisher's exact test).

Despite differences in the abundance and proportions of fusion events observed following nucleofection with 16p/21q TALEN pairs, there was no significant difference ( $p = 0.2039$ ,  $p = 0.3181$ , Fisher's exact tests) in the proportions of '21q', '16p' or 16p:21q inter-chromosomal fusion events between HCT116 WT, *POLQ*<sup>-/-</sup> and *POLQ*<sup>-/-</sup>:*LIG4*<sup>-/-</sup> cell lines (**Figure 4.14B-C**).

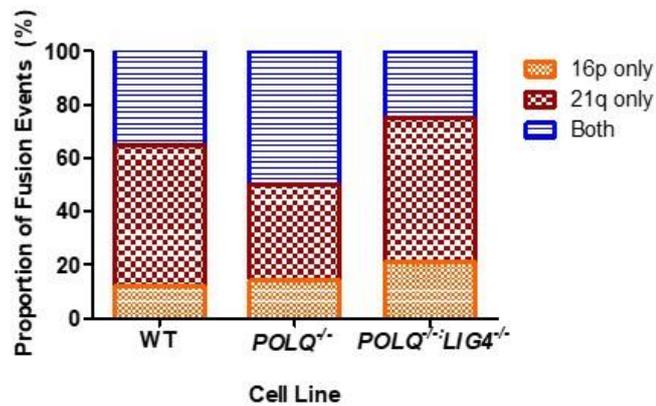
A



B



C



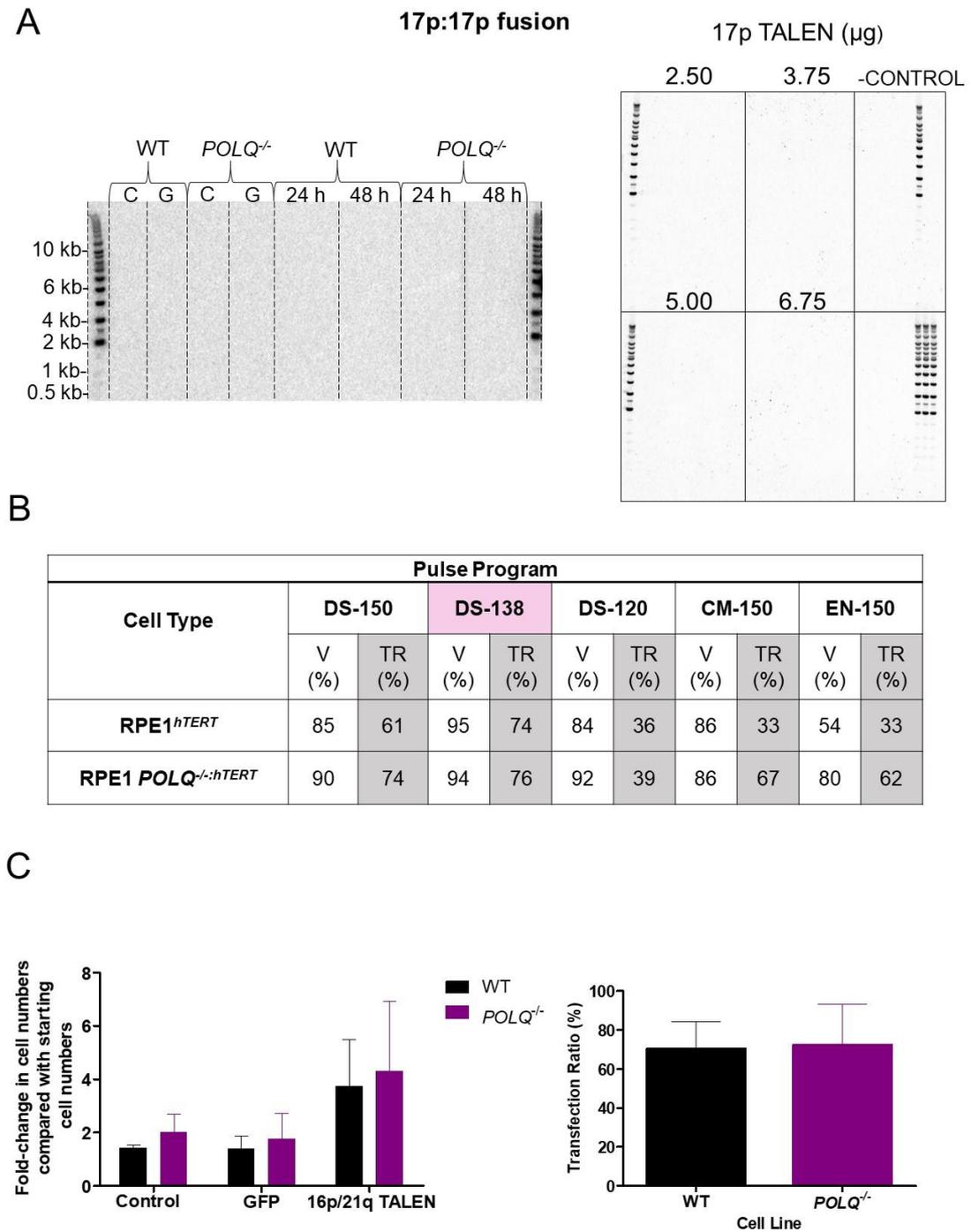
**Figure 4.14:** (A) Fusion blots hybridised with the 16p (top) and 21q probe (bottom). Blots generated following 16p:21q PCR for HCT116 WT,  $POLQ^{-/-}$  and  $POLQ^{-/-}:LIG4^{-/-}$  cell lines 48 h following nucleofection with 16p/21q TALEN pairs in addition to GFP (G) and untransfected controls (C). Fusion frequencies are shown below each sample (B) Bar chart showing the frequency of '16p', '21q' and 16p-21q fusion events following 16p:21q dual-primer PCR for HCT116 WT,  $POLQ^{-/-}$  and  $POLQ^{-/-}:LIG4^{-/-}$  cell lines 48 h following nucleofection with 16p/21q TALEN pairs. Error bars show standard deviation values. (E) Stacked bar chart showing proportions of '16p', '21q' and 16p-21q fusion events following 16p:21q dual-primer PCR detected for HCT116 WT,  $POLQ^{-/-}$  and  $POLQ^{-/-}:LIG4^{-/-}$  cell lines 48 h following nucleofection with 16p/21q TALEN pairs.

#### 4.4.9 Transient Transfection of RPE1<sup>hTERT</sup> cells with TALEN pairs

The RPE1<sup>hTERT</sup> cell line was used to compare DNA repair activity in a non-cancerous cell line with cancerous cell lines (HAP1 and HCT116) in the presence and absence of POLQ. Following nucleofection with 17p TALEN pairs, no 17p telomere fusion events were identified in either WT or POLQ-deficient RPE1<sup>hTERT</sup> cells. This was consistent with previous observations using the 17p TALEN with RPE1<sup>hTERT</sup> cells (Dr Greg Ngo, Cardiff University, personal communication) (**Figure 4.15A**). Therefore, 16p/21q TALEN pairs were utilised, as with the HAP1 cells, to induce DSB in the 16p and 21q family subtelomere sequences.

#### Nucleofection Optimisation in RPE1<sup>hTERT</sup> cells

To determine the optimal conditions for the effective transfection of both RPE1 WT<sup>hTERT</sup> and RPE1 POLQ<sup>-/-hTERT</sup> parental cells, 6 different pulse programs were tested based on their similarity to the DS-150 program previously used by Dr Greg Ngo, Cardiff University. The pulse program DS-138 was chosen as the most appropriate as this offered the highest transfection ratio that was similar for both RPE1 WT<sup>hTERT</sup> and RPE1 POLQ<sup>-/-hTERT</sup> cell lines. This program also did not have much of a detrimental effect on cell viability following transfection. 24 and 48 h following transfection, RPE1<sup>hTERT</sup> cells were counted, the GFP transfection efficiency assay was performed and DNA extracted. 24 and 48 h harvest time points were analysed to determine whether there were any differences in the abundance of telomere fusion events in RPE1<sup>hTERT</sup> cell lines at different time points. The mean GFP transfection efficiency for the RPE1 WT<sup>hTERT</sup> cell line across three experiments was 70.3% compared to a mean GFP transfection efficiency of 72.3% for the RPE1 POLQ<sup>-/-hTERT</sup> parental cell line (**Figure 4.15B-C**).



**Figure 4.15: (A)** 17p single-primer fusion blots for RPE1 WT<sup>hTERT</sup> cells 48 h following nucleofection with 17p TALENs (left) additional blot obtained from Dr Greg Ngo (Cardiff University) to confirm findings (right). **(B)** Table showing cell viability (V) and transfection ratios (TR) for RPE1 WT<sup>hTERT</sup> and RPE1 *POLQ*<sup>-/-hTERT</sup> cell lines 48 h after nucleofection with GFP plasmids using different pulse programs. **(C)** Bar chart showing fold-change in cell numbers 48 h following nucleofection with GFP or 16p:21q TALEN plasmids using the DS-138 pulse program alongside controls compared to starting cell numbers (left) and mean transfection ratios for RPE1 WT<sup>hTERT</sup> and *POLQ*<sup>-/-hTERT</sup> cell lines 48 h following nucleofection with GFP plasmids (right). Data obtained from three replica experiments.

### Abundance of Fusion Events

To analyse any differences in the abundance of telomere fusion events in RPE1<sup>hTERT</sup> cells, 16p and 21q primers were used separately and in combination to detect fusion events involving these telomere families 24 and 48 h following transfection with 16p/21q TALEN pairs.

### Single-primer PCR

No fusion events were detected following single-primer 16p or 21q PCR for either RPE1 WT<sup>hTERT</sup> or *POLQ*<sup>-/-hTERT</sup> cell lines 24 or 48 h after nucleofection with 16p/21q TALEN pairs. There was therefore no change in the abundance of fusion events compared with untransfected or GFP controls compared with cells that had undergone nucleofection with the 16p/21q TALENs.

### 16p:21q dual primer PCR: 24 h harvest

Most fusion events detected 24 h following nucleofection for both the RPE1 WT<sup>hTERT</sup> (63%) and RPE1 *POLQ*<sup>-/-hTERT</sup> cell line (54%) were 16p-21q inter-chromosomal fusion events. There was no statistically significant difference in the proportions of these inter-chromosomal fusion events when comparing the two cell lines ( $p=1.000$ , respectively, Fisher's exact test). Similarly, there was no statistically significant difference in the proportion of '21q' events (WT=25%, *POLQ*<sup>-/-</sup> =46%) or '16p' events (WT=13%, *POLQ*<sup>-/-</sup> =0%) observed 24 h following nucleofection when comparing the cell lines ( $p=0.3810$  and  $p=0.2367$  respectively, Fisher's exact test).

Therefore, for both cell lines, 16p-21q inter-chromosomal fusion events were most abundant overall. However, there was no significant difference in the proportion of 16p:21q inter-chromosomal fusion events compared with '16p' ( $p=0.1189$ , Fisher's exact test) or '21q' ( $p=0.3147$ , Fisher's exact test) events for the RPE1 WT<sup>hTERT</sup> cell line.

As no '16p' events were observed for the RPE1 *POLQ*<sup>-/-:hTERT</sup> cell line, the proportions of 16p-21q inter-chromosomal and '21q' fusion events were significantly greater than the proportion of '16p' events ( $p \leq 0.01$  and  $p < 0.05$ , respectively, Fisher's exact test). **(Figure 4.16).**

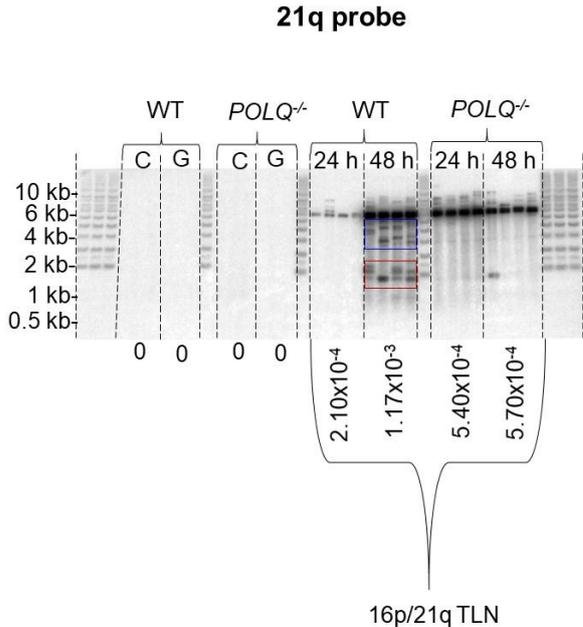
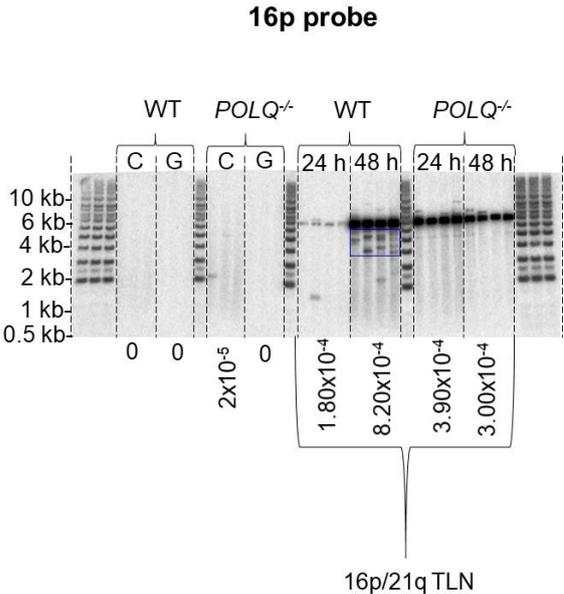
#### **Proportions of 16p:21q fusion events: 48 h harvest**

Most fusion events detected 48 h following nucleofection with the 21q probe were also 16p-21q inter-chromosomal fusion events for the RPE1 *WT*<sup>hTERT</sup> cell line (57%). No '16p' events were detected at this later time point, thus the proportion of '21q' events accounted for 43% of all the 16p:21q events observed. Therefore, a significantly greater proportion of '21q' and 16p:21q inter-chromosomal events were observed compared with '16p' events ( $p \leq 0.001$ , Fisher's exact test). Although a change in the proportion of events detected was observed 48 h compared with 24 h following nucleofection for the RPE1 *WT*<sup>hTERT</sup> cell line, these changes were not statistically significant (16p:21q:  $p = 1.000$ , '21q':  $p = 0.1860$ , '16p':  $p = 0.4462$  respectively, Fisher's exact test).

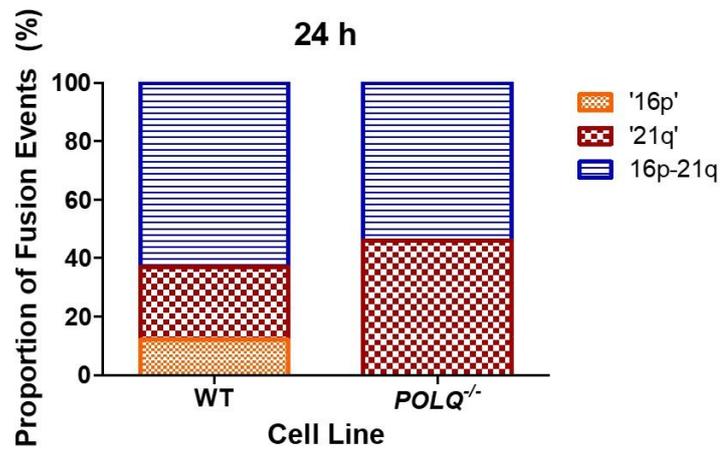
The majority of fusion events detected 48 h following nucleofection for the RPE1 *POLQ*<sup>-/-:hTERT</sup> cell line was '21q' events (68%). As with the 24 h time point, no '16p' only events were observed, therefore 16p-21q inter-chromosomal fusion events accounted for 32% of all 16p-21q events detected. Thus, a significantly greater proportion of 16p-21q and '21q' events were observed compared with '16p' events ( $p < 0.05$  and  $p < 0.01$ , respectively, Fisher's exact test). In addition, there were no significant differences in the proportions of fusion events detected for this cell line at the 24 h time point compared with 48 h time point ( $p = 0.4667$ , Fisher's exact tests).

Overall, no significant differences in the proportions of '16p', '21q' or 16p-21q fusion events 48 h following nucleofection were observed when comparing the two cell lines ( $p = 0.2367$ , Fisher's exact test) **(Figure 4.16).**

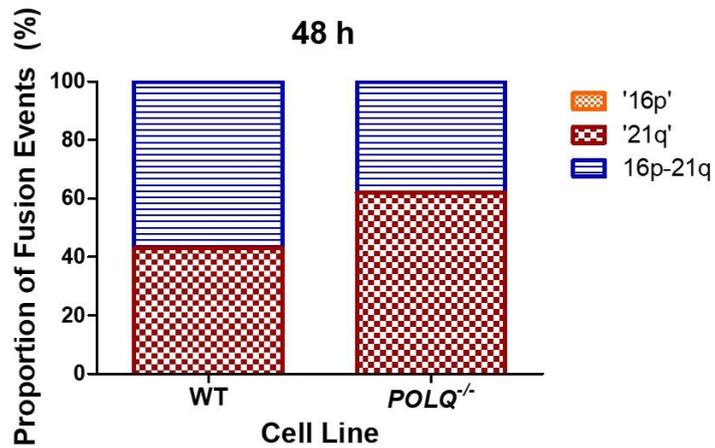
A



B



C



**Figure 4.16: (A)** Fusion blots generated following dual-primer 16p:21q PCR hybridised with the 16p probe (top) and 21q probe (bottom) for RPE1<sup>hTERT</sup> cell lines 24 and 48h following nucleofection with GFP plasmids (G) or 16p:21q TALEN pairs as well as untransfected controls (C). 16p:21q inter-chromosomal fusion events highlighted by blue box, '21q' events highlighted by red box. Fusion frequencies per genome are listed below each sample **(B)** Stacked bar chart showing proportion of '16p', '21q' and 16p:21q fusion events observed following 16p:21q dual-primer PCR for RPE1<sup>hTERT</sup> cell lines 24 h and **(C)** 48 h following nucleofection with 16p:21q TALENs.

## 4.5 DISCUSSION

### 4.5.1 POLQ-deficient cell lines associated with increased inter-chromosomal telomere fusion and inhibition of intra-chromosomal telomere fusion events

For both the HAP1 *POLQ*<sup>-/-:DN-hTERT</sup> and the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cell lines, a significantly greater abundance and proportion of inter-chromosomal fusion events associated with telomere-driven crisis were observed compared with the WT<sup>DN-hTERT</sup> cell lines. In contrast, whilst 17p:17p intra-chromosomal fusion events ( $1.20 \times 10^{-4}$  p/dg) were detected for HCT116 WT<sup>DN-hTERT</sup> clones passaging through telomere-driven crisis, no 17p:17p sister-chromatid fusion events were observed for the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cell line. These data suggest that the disruption of *POLQ* is associated with the inhibition of intra-chromosomal fusion events and an increase in inter-chromosomal fusion events during telomere-driven crisis.

This striking reduction in the frequency of intra-chromosomal fusion events was also confirmed in the HCT116 cell lines using the TALEN system that promotes telomere fusion as a consequence of targeted subtelomeric DSBs. Significantly greater frequencies and proportions of 17p:17p sister chromatid fusion events were detected for the HCT116 WT<sup>DN-hTERT</sup> cell lines compared with the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cell line following nucleofection with 17p TALEN pairs.

The reduction in detectable 17p:17p intra-chromosomal fusion events in the HCT116 *POLQ*-deficient cell line compared with the HCT116 WT line suggests that *POLQ* has an active role in the formation of 17p:17p sister chromatid fusion events. This corroborates previously published data which shows that *POLQ* has a role in A-NHEJ and not C-NHEJ (Kent et al. 2015), therefore, these HAP1 and HCT116 *POLQ*-deficient cell lines are dependent on other forms of DNA repair such as C-NHEJ for facilitating telomere fusion events. Consistent with this is the observation that HCT116 *POLQ*<sup>-/-:LIG4</sup> cells displayed the lowest frequency of fusion events following

nucleofection with TALEN pairs as both A-NHEJ and C-NHEJ are disrupted for this cell line (Lazzerini-Denchi and Sfeir 2016).

16p-21q inter-chromosomal fusion events constituted the most abundant type of fusion event observed for three distinct WT and POLQ-deficient cell lines (HAP1, RPE1<sup>hTERT</sup> and HCT116) following nucleofection with 16p/21q TALEN pairs. However, no significant differences in the abundance or proportions of fusion events were observed for WT cell lines compared with POLQ-deficient cell lines following nucleofection with 16p/21q TALEN pairs.

Taken together, these observations are consistent with a role for POLQ in mediating sister-chromatid fusion. Since POLQ is implicated in A-NHEJ-mediated intra-chromosomal telomere fusion events (Mateos-Gomez et al. 2015) these data conform with the expectation that the absence of POLQ in these cell lines would not impact on the formation of 16p-21q inter-chromosomal fusion events induced following nucleofection with TALEN pairs.

However, the increased abundance of inter-chromosomal fusion events detected for the POLQ-deficient cell lines compared with the WT cell lines following telomere-driven crisis, was not observed following nucleofection with TALEN pairs. As healthy cells are used for nucleofection with TALEN pairs, these data suggest that POLQ does not actively suppress or inhibit C-NHEJ in healthy WT cells. This corroborates previous data that suggests C-NHEJ and A-NHEJ pathways can be employed at the same time and that C-NHEJ can mediate the repair of DSBs within subtelomeric regions (Muraki et al. 2015). The significant increase in the abundance of inter-chromosomal fusion events in both HAP1 *POLQ*<sup>:DN-hTERT</sup> and HCT116 *POLQ*<sup>:-DN-hTERT</sup> clonal populations compared with WT<sup>DN-hTERT</sup> cell lines may be due to the elevated incidence of DSBs that arise during crisis compared with targeted DSBs induced by TALEN pairs. Previously published data suggests that when the incidence of DSBs increases (as with cells going through crisis compared with TALEN-targeted cells),

DNA repair mechanisms that depend on resection such as HR and A-NHEJ are suppressed and an increase in C-NHEJ-mediated repair is observed (Mladenov et al. 2019).

#### 4.5.2 Telomere fusion events during crisis

I have previously shown that the disruption of *POLQ* in both HAP1<sup>DN-hTERT</sup> and HCT116<sup>DN-hTERT</sup> cells resulted in these cell lines escaping crisis more readily than the WT cell lines (*Chapter 3*). This was defined by a more rapid return to pre-crisis growth rates in the HAP1 *POLQ*<sup>-/-:DN-hTERT</sup> compared with the HAP1 WT<sup>DN-hTERT</sup> cell line, whereas no reduction in growth rate associated with crisis was observed for the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cell line. It is possible that the differences in the type of fusion events observed has contributed to the differences in the ability of these clonal populations for escaping crisis.

The reduction of intra-chromosomal telomere fusion events in the *POLQ*-deficient cell lines going through crisis reflects previously published data that demonstrated that *POLQ* is a component of the A-NHEJ pathway (Mateos-Gomez et al. 2015) which gives rise to intra-chromosomal fusion events during telomere-driven crisis (Jones et al. 2014). Therefore, the disruption of *POLQ* in these cell lines resulted in abrogation of the A-NHEJ pathway and a decrease in the formation of intra-chromosomal telomere fusion events. The suppression of A-NHEJ in these cells did not prevent the escape from crisis but was associated with a more timely escape from crisis compared with WT cell lines.

Telomere-driven crisis is usually driven by critical telomere shortening and is associated with an increasing incidence of telomere fusions. Both HAP1 *POLQ*<sup>-DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cell lines had longer starting telomere lengths than the HAP1 WT<sup>DN-hTERT</sup> and HCT116 WT<sup>DN-hTERT</sup> cell lines. However, the telomere lengths measured for the HAP1 *POLQ*<sup>-DN-hTERT</sup> clones did not reach previously recognised thresholds for the onset of crisis (Baird et al. 2003). At these longer

telomere lengths, it is likely that there was no reduction in telomere function as TRF2 binding is not inhibited, allowing T-loop formation. It is possible that telomere-driven crisis was initiated for this cell line by the accumulation of a few critically short telomeres (Counter et al. 1992) at other chromosome ends that were not examined in this study. However, the telomere fusion events observed for this cell line may be due to enhanced replication stress or fork collapse as telomere repeat arrays are late-replicating sequences (Arnoult et al. 2010) within these rapidly-dividing cells. Stalled replication forks can collapse into DSBs (Muraki and Murnane 2017) which then become substrates for DSB repair leading to the formation of telomere fusion events.

The increase in inter-chromosomal telomere fusion events in the POLQ-deficient cell lines going through crisis may suggest that their increased capacity for crisis escape may be influenced by their ability to tolerate C-NHEJ-mediated recombinations. However, previous data has shown that escape from crisis is dependent on LIG3 (Jones et al. 2014), which is involved in ligating DSBs via the A-NHEJ pathway. This study demonstrated that LIG3-deficient clonal populations of HCT116<sup>DN-hTERT</sup> cells failed to escape crisis and had increased levels of C-NHEJ-mediated inter-chromosomal fusion events. In contrast, LIG4-deficient and WT HCT116<sup>DN-hTERT</sup> cells that escaped crisis had an increased abundance of intra-chromosomal telomere fusion events.

It has been shown that the POLQ helicase domain is able to actively displace both RAD51 (Ceccaldi et al. 2015) and Replication Protein A (RPA) (Mateos-Gomez et al. 2017) from ssDNA. It is therefore possible that the disruption of *POLQ* not only results in the suppression of A-NHEJ but may also cause the upregulation of HR, SSA and BIR as RPA and RAD51 (components in these repair mechanisms) are no longer antagonised by POLQ. Therefore, POLQ-deficient cells are able to employ a variety of repair mechanisms following DNA damage. For LIG3-deficient cells in which A-NHEJ is suppressed, POLQ-helicase activity may result in the inhibition of other repair

pathways. This would mean that LIG3-deficient cells are dependent on C-NHEJ and HR for DSB repair. However, if resection by MRN-CtIP occurs, repair is dependent on either HR, A-NHEJ or SSA and cannot be salvaged by C-NHEJ (Shamanna et al. 2016). If this repair occurs during S/G<sub>2</sub>-phases of the cell cycle, then A-NHEJ mediated via LIG1 may contribute to repair (Liddiard et al. 2018). However, unresolved DSBs that cannot be repaired by LIG1-mediated A-NHEJ or C-NHEJ could explain the inability of LIG3-deficient cells to escape crisis (Jones et al. 2014).

In addition, as LIG3 is required for BER, any oxidative damage that may have occurred due to crisis-induced reactive oxygen species (ROS) (Fouquerel et al. 2019) might remain unresolved. The mounting levels of unresolved DNA damage may therefore result in cell death, providing a potential explanation for the failure of LIG3-deficient cells to escape crisis.

### 4.5.3 Conclusions

Data presented in this chapter corroborates previous data which demonstrates a role for POLQ in mediating sister-chromatid telomere fusion via A-NHEJ. The disruption of *POLQ* was associated with an inhibition of intra-chromosomal and an increase in inter-chromosomal telomere fusion events during telomere-driven crisis. These observations were made for two distinct cell lines passaging through telomere-driven crisis. Likewise, a reduction in intra-chromosomal telomere fusion events was also detected following DSB repair within subtelomeric DNA. Telomere fusion events mediated by A-NHEJ and C-NHEJ may impact on the ability of cells to escape telomere-driven crisis. These data suggest that the disruption of *POLQ* during crisis results in the upregulation of C-NHEJ-mediated repair which may therefore be associated with recombinations that cause more effective reactivation of telomere maintenance mechanisms.

*POLQ* overexpression has been observed in a variety of human cancers and linked to reduced breast cancer patient survival. However, data presented here may inform

whether clinical POLQ inhibitors represent valid therapeutic targets as clinical POLQ inhibition may drive tumoural heterogeneity and survival.

## Chapter 5

### How does the loss of POLQ impact on the evolving cancer genome?

#### 5.1 ABSTRACT

HAP1 *POLQ*<sup>:DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones escaped telomere-driven crisis more readily than their WT counterparts and showed different spectra of telomere fusion events. To gain an understanding of the underlying mechanisms of POLQ-mediated telomere fusion, fusion amplicons from HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>:DN-hTERT</sup> clones were sequenced using Illumina HiSeq4000 paired-end high-throughput sequencing. Fusion events were categorised as genomic fusion events, 16p-21q inter-chromosomal fusion events or '16p' / '21q' fusion events. A significantly greater proportion of '16p' / '21q' events (80%) and a significantly smaller proportion of both genomic (18%) and inter-chromosomal (2%) fusion events were observed for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line.

Furthermore, in order to establish the key mutational events that accompany the escape from crisis in the absence of POLQ, whole genome sequencing was performed for HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>:DN-hTERT</sup> clones before and after the escape from crisis. From these data I was able to examine how a telomere-driven crisis in the presence and absence of POLQ affects genomic stability and rearrangements at non-telomeric loci. An increased proportion of HAP1 *POLQ*<sup>:DN-hTERT</sup> clones were polyclonal following crisis (75% compared to 25%) suggesting that the increased abundance of telomere fusion events gave rise to increased genomic heterogeneity. This could be linked to the ability of these cells to escape crisis more readily than the WT cell line. Altogether, these data suggest that the increased abundance of inter-chromosomal telomere fusion events, including fusion events that incorporate genomic loci may facilitate the escape from telomere-driven crisis.

## 5.2 INTRODUCTION

It has previously been shown that POLQ is overexpressed in several human cancers and that this overexpression was correlated with poor clinical outcome. In addition, the overexpression of POLQ was associated with gross chromosomal changes including dicentric chromosomes and end-to-end fusions (Lemée et al. 2010). These gross chromosomal rearrangements are common during carcinogenesis (Mitelman et al. 1997) and may be initiated by the inappropriate repair of critically short telomeres. The activation of telomerase in these cells is coincident with escape from telomere crisis via genome stabilisation and therefore advancing malignancy.

I have previously shown that both HAP1<sup>DN-hTERT</sup> and HCT116<sup>DN-hTERT</sup> POLQ-deficient cell lines escaped crisis more readily than their WT counterparts (*Chapter 3*). This may have been influenced by the type of telomere fusion events that occurred during crisis in the presence / absence of POLQ. Different spectra of fusion events were observed following fusion PCR for the POLQ-deficient cell lines compared with the WT cell lines (*Chapter 4*). This is consistent with previous data which demonstrates POLQ is a component of the A-NHEJ pathway (Mateos-Gomez et al. 2015) that is associated with intra-chromosomal fusion events whereas C-NHEJ does not require POLQ and mediates inter-chromosomal fusion events (Jones et al. 2014). However, these previous data also suggest that increased proportions of inter-chromosomal fusion events (as would be expected in POLQ-deficient cell lines) failed to facilitate the escape from crisis.

Therefore, the characterisation of telomere fusion events observed for both WT and POLQ-deficient cell lines is required. In addition, whole genome sequencing analysis of these cell lines allows for comparisons between fusion profiles and genome-wide stability.

### 5.3 AIMS OF THE CHAPTER

Different spectra of fusion events were observed for WT and POLQ-deficient cell lines as they transited telomere-driven crisis. The purpose of this chapter was to provide a detailed characterisation of the telomere fusion events that arose during telomere-driven crisis in HAP1 WT and HAP1 *POLQ*<sup>-</sup> cell lines. This was done using Illumina HiSeq4000 paired-end next-generation sequencing of telomere fusion amplicons from HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>-DN-hTERT</sup> cells going through telomere-driven crisis. An additional objective was to analyse the impact of telomere fusion events that arose during crisis on genome stability in HAP1 WT and HAP1 *POLQ*<sup>-</sup> cell lines. This was achieved by performing whole genome sequencing for HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>-DN-hTERT</sup> clones before and after crisis.

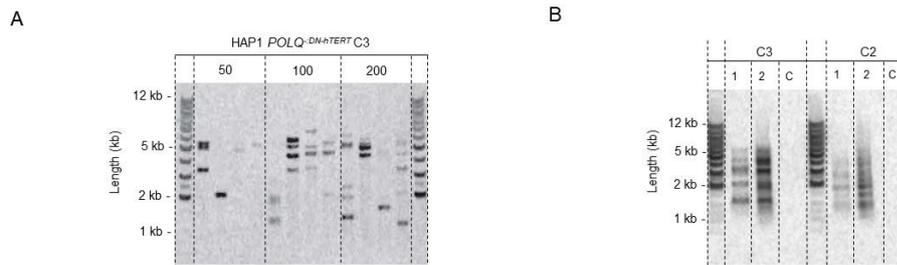
The aims of this chapter were as follows:

- To investigate if there were any differences in the overall abundance and the type of fusion events ('16p', '21q', 16p-21q) in the absence of POLQ
- To establish if there were any differences in processing of fusion events in terms of microhomology usage and insertions at fusion junctions
- To determine differences in the abundance and type of structural and copy number variations following crisis in the absence of POLQ
- To establish if there were any differences in the clonality of the cell lines before and after crisis
- To evaluate whether there were similarities in any genomic loci incorporated into fusion junctions for the two cell lines
- To assess whether genomic telomere fusion events correlated with any structural variants detected following crisis

## 5.4 Results

### 5.4.1 Sequence analysis of HAP1<sup>DN-hTERT</sup> fusion amplicons generated during telomere-driven crisis

To further examine any differences in the abundance or type ('16p', '21q' or 16p-21q) of telomere fusion events that arose during telomere-driven crisis, sequence analysis of telomere fusion amplicons was performed for both HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>DN-hTERT</sup> clones (*Chapter 3*). For each cell line, 5 clones with the highest fusion frequency (associated with telomere-driven crisis) were selected for Illumina HiSeq4000 paired-end high-throughput sequencing of telomere fusion amplicons. Telomere fusions are detectable using this method as the distance between each paired-read is known. Moreover, this method ensures repetitive regions (such as telomeric DNA) can be sequenced with high mapping quality (*Chapter 2*). To maximise the number of fusion events amplified for sequencing, the 16p:21q fusion PCR was first optimised by testing three different genomic DNA (gDNA) input amounts: 50, 100 and 200 ng (**Figure 5.1A**). Subsequently, 96 reactions using 200 ng of DNA per reaction were performed for each sample and then purified using Agencourt AMPure XP beads (Beckman Coulter) (**Figure 5.1B**). Whilst DNA input amounts were comparable for each sample, this did not result in equivalent amplicon yields for sequencing, owing to the inherent differences in fusion frequency amongst samples. PCR products were verified by Southern blotting before and after purification to check that there were no significant losses or changes in MW distributions of amplicons. Samples were then sent for paired-end Illumina HiSeq4000 PE100 amplicon sequencing at the Wales Gene Park.



**Figure 5.1:** (A) 16p:21q fusion assay for HAP1 *POLQ-DN-hTERT* Clone 3 (C3) using 3 different gDNA input amounts: 50, 100 and 200 ng. Telomere fusions were detected using the 21q probe. (B) Pooled fusion PCR products following 96 reactions with a concentration of 200 ng of gDNA per reaction. (1) PCR products before and (2) after purification for C3 and HAP1 *POLQ-DN-hTERT* Clone 2 (C2) in addition to negative control (H<sub>2</sub>O) (C) (5% of the total volume was loaded for each sample). Fusion events were detected with 21q probe.

#### 5.4.2 Mapping pipeline

Following high-throughput sequencing, paired-end read data for each sample was processed for mapping genomic, 16p-21q family inter-chromosomal and '16p' / '21q' family telomere fusion events. Sequence tags and primer sequences were removed from the reads, which were subsequently mapped to the GRCh38 reference genome. The mapping and filtering of read-pairs was performed by Dr Kez Cleal, Cardiff University. BAM and BED files were produced for visualisation of data spread along the subtelomeric sequence in the Integrative Genomics Browser (IGV) (Robinson et al. 2011) and read-pair mapping coordinates were listed in excel files.

#### 5.4.3 Classification of events identified

All consensus reads recorded were visualised using IGV and BLAST-alignments were performed to validate genuine telomere fusion events. To identify high quality events and remove potentially false positive events, fusions were further classified using a machine learning approach. Briefly, a large training dataset was prepared by Dr Kate Liddiard (Cardiff University) that consisted of 20,000 candidate calls across 24 separate samples in fibroblast cell lines manually evaluated as true or false positive fusion events. A random forest model was then trained on these data

achieving a cross-validated precision of 0.8 and recall of 0.75. Therefore, a probability value of 0.5 would be associated with a precision of 0.8 and recall of 0.75.

All telomere fusion events were further classified depending on the locations involved in the fusion event. Events involving subtelomeric DNA and other parts of the genome are referred to as genomic fusion events. '16p-21q inter-chromosomal' fusion events were inferred from read-pairs that linked these two distinct subtelomeric families. '16p' or '21q' fusions consist of read-pairs that mapped to one family only. However, it was not possible to determine the precise family members involved.

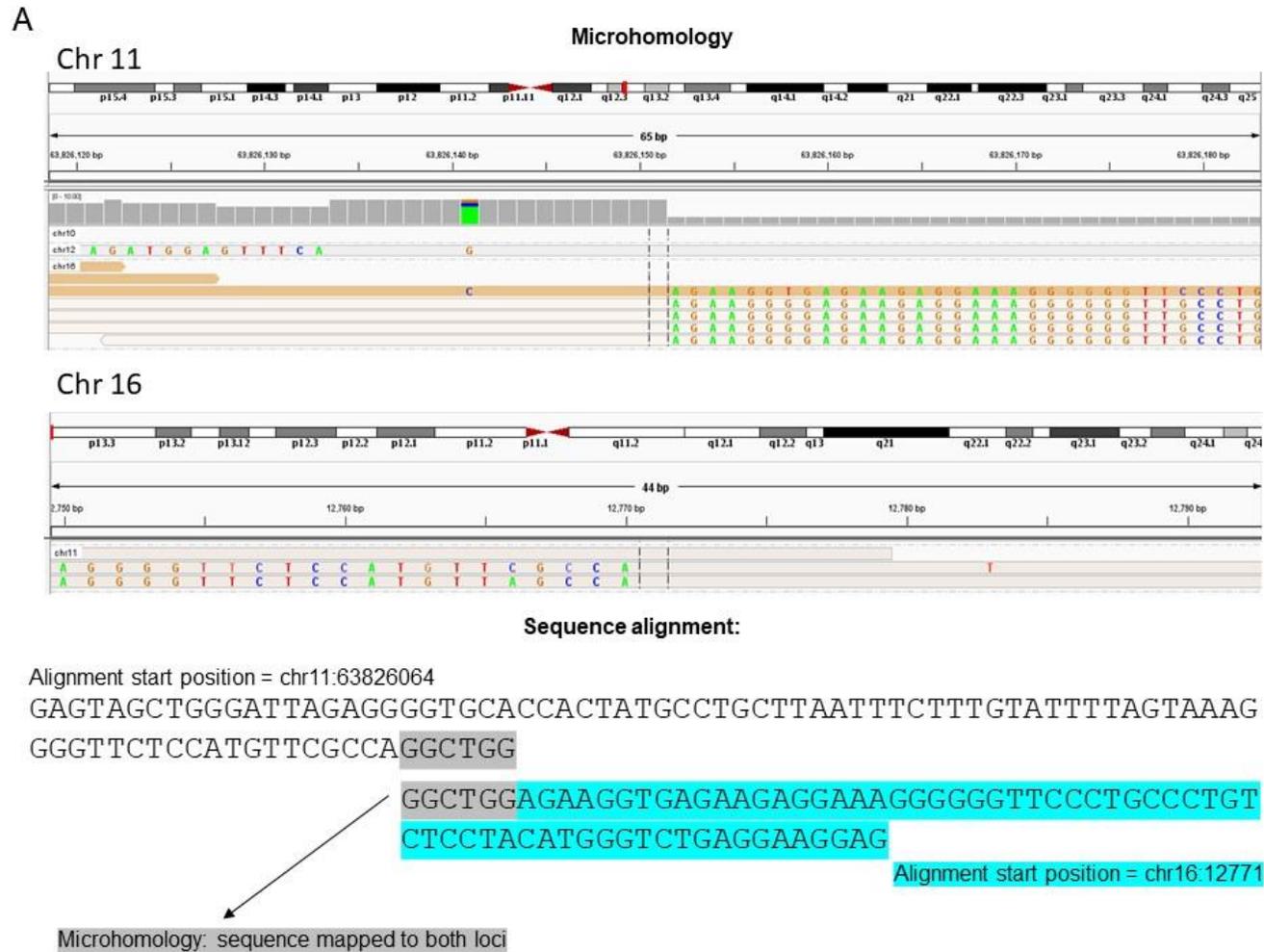
#### **5.4.4 Manual validation of fusion events**

To further validate the read-pairs classified using the machine-learning approach, all paired-reads and their coordinates were recorded in an excel file for manual curation and sequence verification for each sample. IGV was used to visualise read-pairs before alignment to the human genome (Ensembl BLAST) for sequence verification. This enabled certification that the mapping corresponded to a unique and accurate location and confirmed the event was in the expected orientation (subtelomere-genomic location or subtelomere-subtelomere).

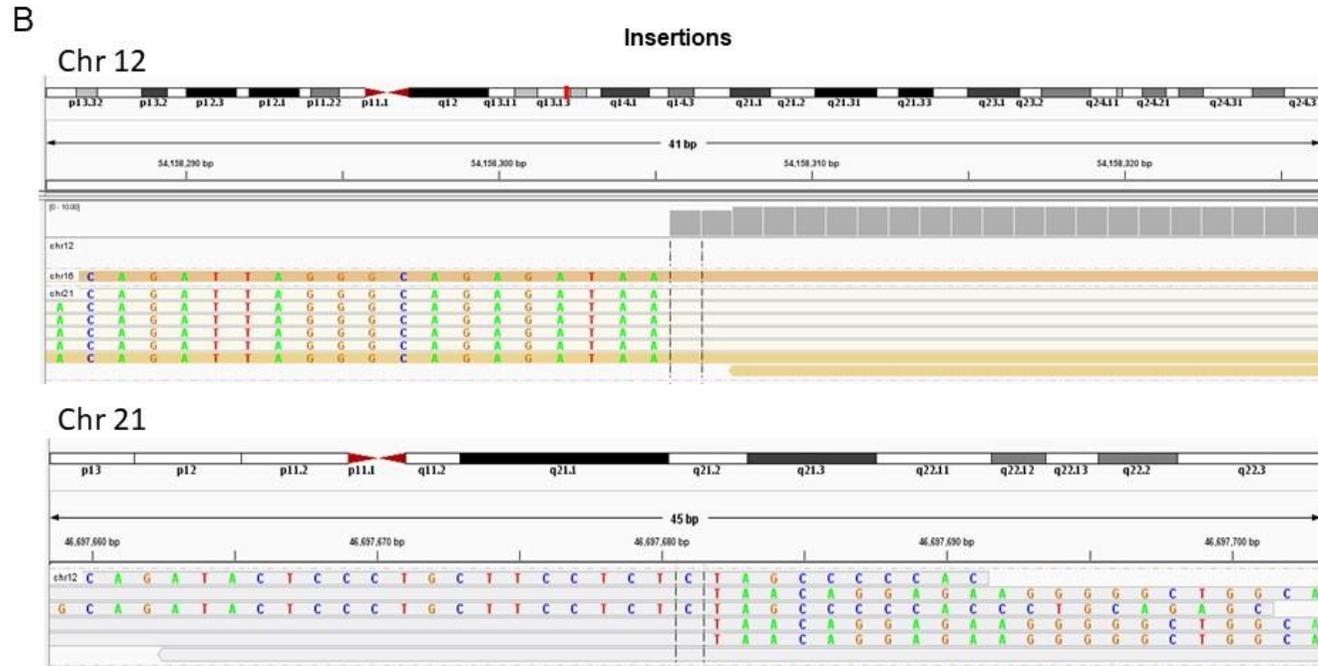
#### **5.4.5 Genomic Fusion Events**

For genomic fusion events, reads were visualised using IGV and BLAST-alignment was used to validate that the events were genuine telomere fusion events, that is, a fusion event between the 16p or 21q subtelomere and another part of the genome. In addition, whether fusion events overlapped with genes was recorded to determine the potential impact of these telomere fusion events on the genome more widely. Furthermore, to assess whether recognised chromatin states or features are linked to telomere fusion, the involvement of CpG islands (regions with high frequencies of CpG dinucleotides) (Guo et al. 2017) and fragile sites, which are prone to genomic

instability, (Durkin and Glover 2007) were recorded. As it has been demonstrated that POLQ is often responsible for insertions and microhomology at DSB sites, any microhomology and insertions at the fusion junctions were also documented. For the following analysis microhomology was defined as 1 bp or more of sequence overlap between fusion components that could be contributed to either component. **(Figure 5.2A)**. Insertions were defined as sequences that could not be aligned to either of the fusion components **(Figure 5.2B)**.



**Figure 5.2: (A)** Example of how microhomology is identified at fusion junctions using IGV and BLAST alignment (Ensembl).



Alignment start position = chr12:54158306

```
ATCATCATTTACTCTGCAATAAAAATGTCAGAATATTGCTAGCTACAT
GCTGATAAGAAGGACAAAGGGGACATTTTTAAGAGAAACCTGGCACCA
TAAGTACAGATTAGGGCAGAGATAA
```

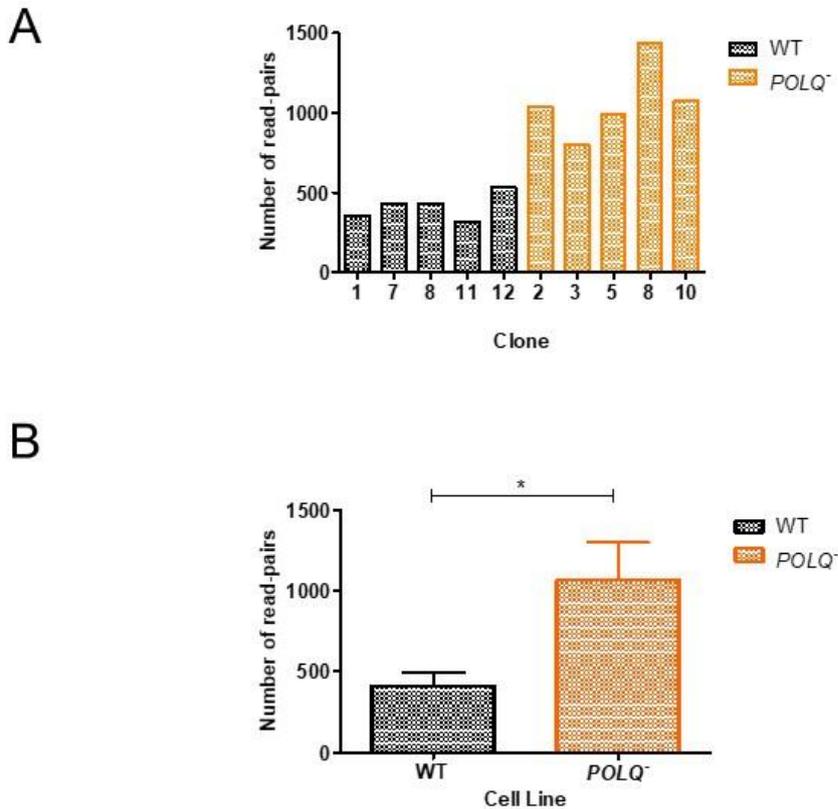
```
TAA CAGGAGAAGGGGGCTGGCAGCCAAAATGCCTGGTGCTA
CTCACCA Alignment start position = chr21:46697579
```

**Insertion: sequence not mapped to either loci**

**Figure 5.2: (B)** Example of how insertions are identified at fusion junctions using IGV and BLAST alignment (Ensembl).

#### 5.4.6 Sequencing Output

A significant 2.6-fold greater number of read-pairs was observed for the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line compared with the HAP1 WT<sup>DN-hTERT</sup> cell line ( $p < 0.05$ , Mann Whitney test) (all read-pairs). For the HAP1 WT<sup>DN-hTERT</sup> cell line, the read-pair yields per sample ranged from 318 to 531. Overall, 2049 read-pairs were recorded for this cell line. In contrast, for the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line, the read-pair yields per sample ranged from 801 to 1440. Overall, 5342 reads were recorded for this cell line (**Figure 5.3A-B**). As these differences were consistent amongst the cell lines, (that is, for all 5 HAP1 WT<sup>DN-hTERT</sup> clones, a smaller number of reads were recorded compared with the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line), this observation seems to reflect a genuine difference in the abundance of fusion events between the two cell lines. However, this did not reflect the overall fusion frequency calculated based on events detected by Southern blotting following 16p:21q fusion PCR (*Chapter 4*). This previous analysis showed no significant difference in the overall frequency of fusion events between the two cell lines ( $p = 0.9768$ , Mann Whitney test). Alternatively, this disparity may be due to a bias in extracting higher MW DNA from the *POLQ*-deficient cells. Due to the vast number of read-pairs generated and the limited time available, 100 read-pairs per clone were manually validated for all event types (based on likelihood of true event, determined by machine learning program). As only 18 read-pairs were recorded for the 16p-21q events in the HAP1 WT<sup>DN-hTERT</sup> cell line, this validation is equivalent to 50% of the total read-pairs for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with 28% of the total read-pairs for the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line. These validated events were used for the analysis presented in this chapter.



**Figure 5.3:** (A) Bar chart showing total read-pair yields per clone for HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-DN-hTERT</sup> cell lines following sequencing of fusion amplicons. (B) Bar chart showing the mean number of read-pair yields per HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-DN-hTERT</sup> cell lines following sequencing of fusion amplicons. Error bars indicate standard deviation and significance was determined using a Mann Whitney test.

#### 5.4.7 Genomic Events

Consistent with the elevated overall incidence of read-pairs mapping telomere fusions in the HAP1 POLQ<sup>-DN-hTERT</sup> cell line, the total number of manually-validated genomic fusion events was also higher for the HAP1 POLQ<sup>-DN-hTERT</sup> cell line compared with the HAP1 WT<sup>DN-hTERT</sup> cell line. For the HAP1 WT<sup>DN-hTERT</sup> clones, the total number of valid genomic events (out the total of 500 validated events) recorded was 87 compared to 350 for the HAP1 POLQ<sup>-DN-hTERT</sup> clones (**Figure 5.4A**). The range in the number of fusion events recorded per clone was between 2 and 30 events for HAP1 WT<sup>DN-hTERT</sup> clones compared with 43 to 96 events per HAP1 POLQ<sup>-DN-hTERT</sup> clone (**Figure 5.4A**). Therefore, a significant difference in the abundance of genomic fusion events was observed for POLQ-deficient cells compared with WT cells ( $p \leq 0.01$ , Mann Whitney

test) (**Figure 5.4B**). Previous analysis of Southern blots following 16p:21q fusion PCR showed a significantly greater frequency of 16p-21q inter-chromosomal fusion events for the HAP1 *POLQ*<sup>-DN-hTERT</sup> compared with the HAP1 WT<sup>DN-hTERT</sup> cell line. Therefore, as these genomic telomere fusion events are also inter-chromosomal telomere fusion events, this discrepancy in the abundance of genomic events reflects what was previously observed (*Chapter 4*).

Fusion junctions were analysed and categorised as either blunt (no microhomology or insertions), having 1 or more bp of microhomology or having insertions. For the HAP1 WT<sup>DN-hTERT</sup> clones, fusion junctions with microhomology were the most common (54.7%), although there was no significant difference in the proportion of these junctions and blunt fusion junctions (40.7%) for this cell line. However, junctions with insertions (4.5%) were significantly less common than blunt fusion junctions and fusion junctions with microhomology ( $p \leq 0.001$ , Fisher's exact test).

Similarly, fusion junctions with microhomology was the most common fusion junction feature observed for genomic telomere fusion events for the HAP1 *POLQ*<sup>-DN-hTERT</sup> clones (57.5%). However, there was a significantly greater proportion of junctions with microhomology compared to blunt junctions (32.8%,  $p \leq 0.001$ , Fisher's exact test). There were significantly more blunt fusion junctions and fusion junctions with microhomology than fusion junctions with insertions (9.4%,  $p \leq 0.001$ , Fisher's exact test) (**Figure 5.4B**).

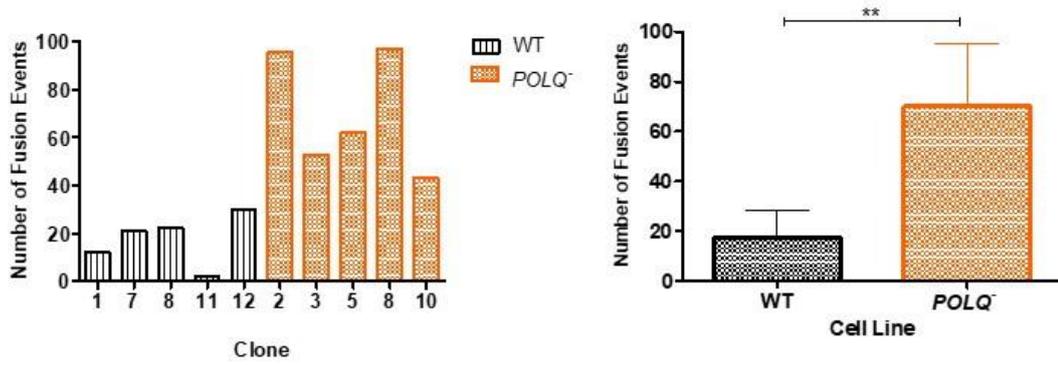
Despite differences in the number of genomic telomere fusion events for the HAP1 WT compared to the HAP1 *POLQ*<sup>-</sup> cell line, there were no significant differences in the proportion of fusion junction features (insertions, blunt, microhomology) when comparing the two cell lines (**Figure 5.4C**). However, the length of the microhomology (WT mean 7.95 bp and *POLQ*<sup>-</sup> mean 3.50 bp) and the insertions (WT mean 21.8 bp and *POLQ*<sup>-</sup> mean 7.09 bp) were significantly longer in the HAP1 WT<sup>DN-hTERT</sup> compared to the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line ( $p \leq 0.001$  and  $p < 0.05$  respectively, Mann Whitney test). In addition, the distributions of microhomology and insertion lengths were varied

when comparing the cell lines as the HAP1 *POLQ<sup>-DN-hTERT</sup>* cell line had much tighter distributions with a few outliers. This is apparent when comparing the differences in the standard deviation values calculated for the distribution of microhomology lengths for the HAP1 *WT<sup>DN-hTERT</sup>* clones (8.75 bp) compared with the HAP1 *POLQ<sup>-DN-hTERT</sup>* clones (4.10 bp). Similarly, the difference in the distribution of insertion lengths is demonstrated by comparing the standard deviation values calculated for insertion length distributions for HAP1 *WT<sup>DN-hTERT</sup>* clones (17.6 bp) with the HAP1 *POLQ<sup>-DN-hTERT</sup>* clones (10.9 bp) (**Figure 5.4D**).

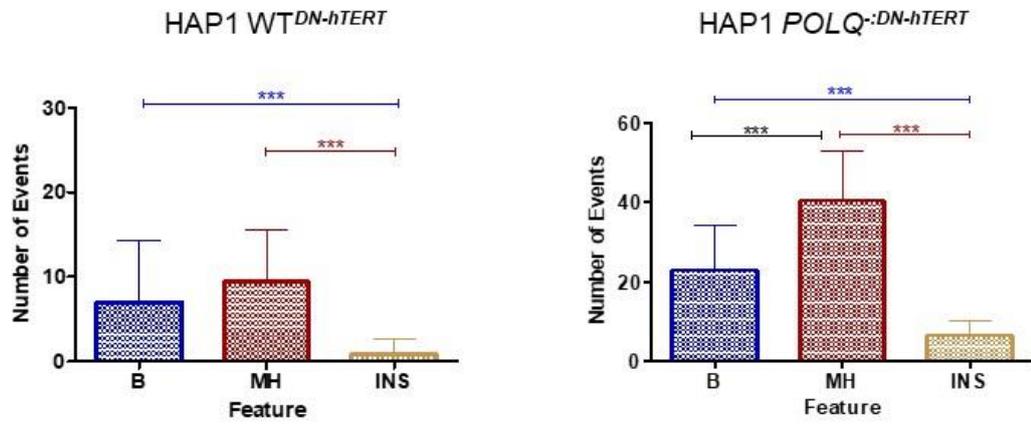
Insertions  $\geq 3$  bp were analysed to determine whether they were templated or non-templated. Insertion sequences were categorised as templated insertions if the exact insertion sequence was mapped to within 100 bp of the fusion junction. For the HAP1 *WT<sup>DN-hTERT</sup>* clones, only 4 genomic fusion events with insertions at the fusion junction were identified. 3 of these events (75%) were templated insertions whereas 1 insertion was a non-templated insertion (25%). Insertions varied in length from 4 bp to 44 bp.

For the HAP1 *POLQ<sup>-DN-hTERT</sup>* clones, 32 genomic fusion events with insertions at the fusion junction were identified. 15 of these events (46.9%) were  $\leq 2$  bp and therefore were categorised as 'undetermined'. For insertions  $\geq 3$  bp, 7 events (21.9%) were templated and insertion length varied from 3 to 33 bp. 10 events (31.3%) were non-templated insertions and the lengths recorded were between 3 to 30 bp. Despite differences in the length distribution of the insertions, there were no significant differences in the proportions of templated and non-templated insertions when comparing the two cell lines ( $p=0.3108$ , Fisher's exact tests) (**Figure 5.4E**).

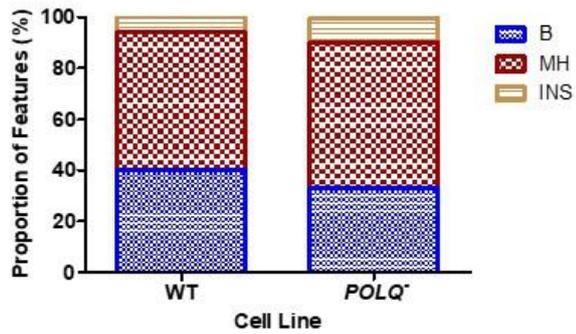
A

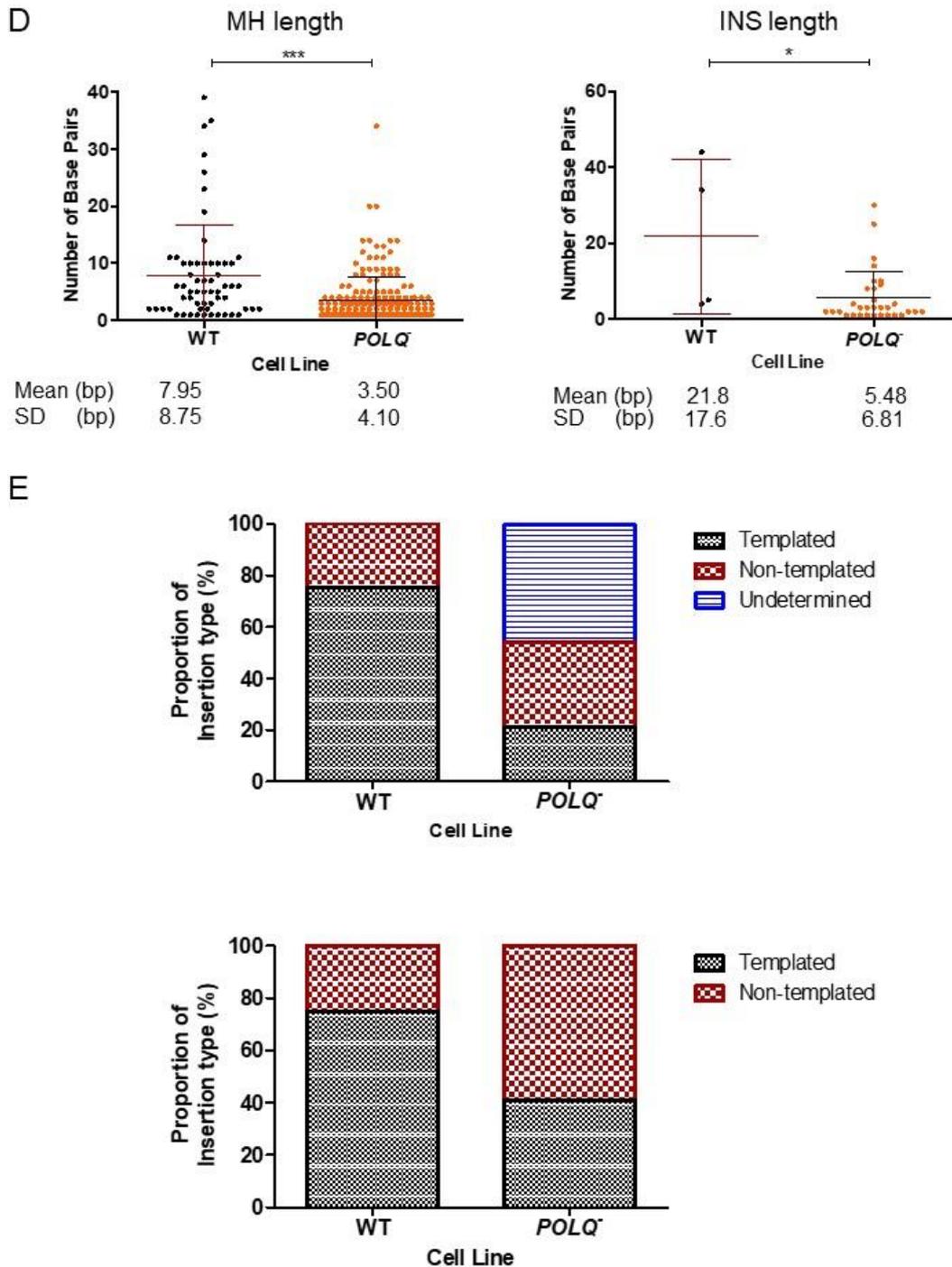


B



C





**Figure 5.4:** (A) Charts showing the total number of validated genomic fusion events for HAP1<sup>DN-hTERT</sup> cell lines. Error bars indicate standard deviation values and significance was determined using a Mann Whitney test. (B) Bar chart showing the mean number of fusion junction features (Blunt = B, Microhomology = MH and Insertions = INS) for HAP1 WT and HAP1 *POLQ* cell lines. Error bars indicate standard deviation values and significance was determined using a Mann Whitney test. (C) Stacked bar chart showing no significant difference in the proportion of fusion junction features for HAP1 WT and HAP1 *POLQ* cell lines. Significance was determined using a Fisher's exact test. (D) Scatter plot showing length of microhomology (left) and insertions (right) at fusion junctions. Error bars indicate mean and standard deviation values and significance was determined using a Mann Whitney test. (E) Stacked bar charts showing the proportions of undetermined, (top) templated and non-templated (bottom) insertions at fusion junctions for HAP1 WT and HAP1 *POLQ* cell lines. Significance was determined using a Fisher's exact test.

### Genomic Locations

The genomic location of each 16p- or 21q- linked fusion event was recorded for both HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-DN-hTERT</sup> clones to compare any differences in chromosomes involved in the fusion events. For both cell lines, the greatest abundance of fusion events involved chromosome 1, likely due to its size as chromosome 1 is the largest chromosome. In addition, no fusion events involved chromosome 21 for either cell line. Interestingly, no fusion events involved chromosome 10 for the HAP1 WT<sup>DN-hTERT</sup> cell line, however, fusion events involving chromosome 10 were observed for the HAP1 POLQ<sup>-DN-hTERT</sup> cell line. **(Figure 5.5A-B).**

The frequency of fusion events per chromosome was analysed based on chromosome size to determine whether there were any chromosomes that had a particular abundance or deficiency of fusion events (*Chapter 2*). The mean fusion frequency used for the basis for these predictions was  $2.28 \times 10^{-10}$  and  $1.16 \times 10^{-4}$  for the HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-DN-hTERT</sup> cell lines respectively. This analysis was based on the manually validated events.

Interestingly, for both cell lines, a decreased fusion frequency was observed for larger chromosomes. For the HAP1 WT<sup>DN-hTERT</sup> cell line, chromosomes 1 – 10 (excluding chromosomes 8 and 9) were involved in fewer fusion events than predicted based on chromosome size. Similarly, for the HAP1 POLQ<sup>-DN-hTERT</sup> cell line, chromosomes 1 – 10 were likewise involved in fewer fusion event than anticipated based on individual chromosome size. Thus, a reduction in fusion frequency seems to be associated with chromosome size as chromosomes 1 – 10 are all  $\geq 133797$  kb **(Figure 5.5C).**

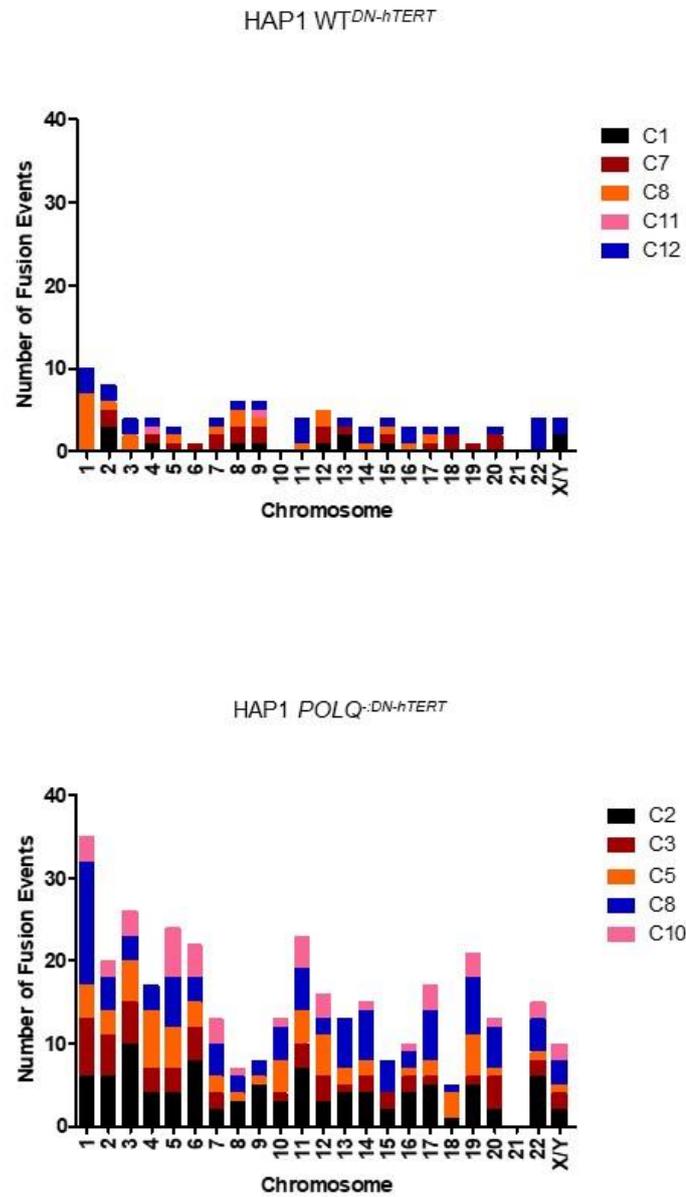
Conversely, the smaller chromosomes were involved in a greater number of fusion events than predicted based on chromosome size for both cell lines. For the HAP1 WT<sup>DN-hTERT</sup> cell line, chromosomes 15 – 22 (excluding chromosomes 19 and 21) were involved in more fusion events than expected based on chromosome size. These

chromosomes are between 50818 and 101991 kb whereas a deficit in fusion events was seen in larger chromosomes which were  $\geq 133797$  kb. The greatest discrepancy between predicted fusion frequency and the actual fusion frequency observed for this cell line involved chromosome 22. A 6.8-fold greater abundance of fusion events involving this chromosome was recorded compared with the predicted frequency based on chromosome size. However, these fusion events were only observed for 1 of the 5 clones (C12) (**Figure 5.5C**).

Likewise, for the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line, chromosomes 14, 19, 20 and 22 were involved in a greater number of fusion events than expected based on their size. These chromosomes range in size from 50818 to 107044 kb. A 6.37, 3.26 and 6.05-fold greater number of fusion events involving chromosome 19, 20 and 22 respectively was observed compared with the predicted fusion frequencies based on chromosome sizes. Therefore, for both cell lines, increased fusion frequency seems to be associated with smaller chromosomes except for chromosome 21 which is the smallest human chromosome at 46710 kb (**Figure 5.5C**).

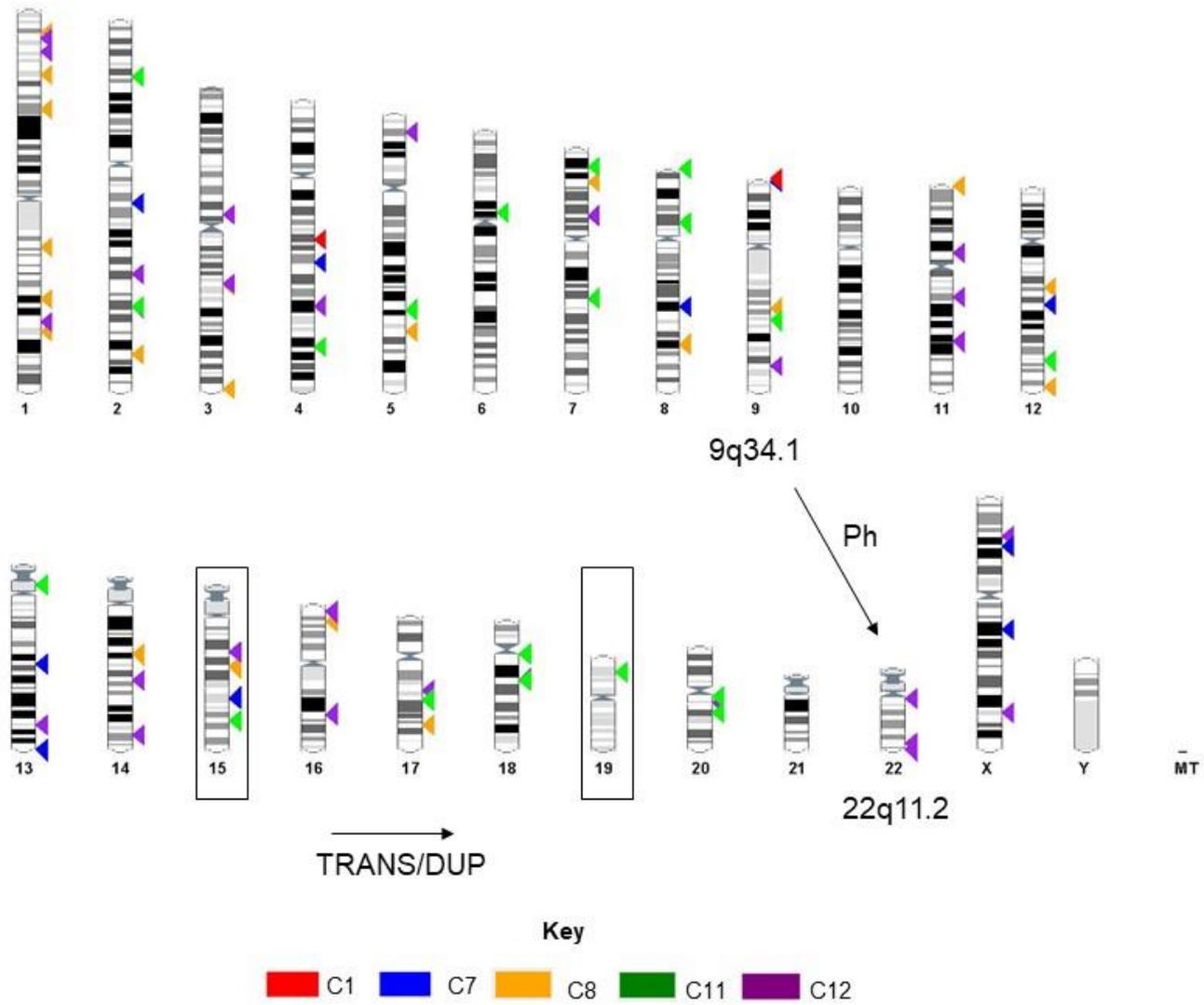
HAP1 cells are derived from the KBM7 CML line that have the Philadelphia translocation. This reciprocal translocation affects chromosome 22 and is caused by a rearrangement of genetic material between chromosome 9 and chromosome 22 (Nowell and Hungerford 1960). As a result, chromosome 22 is defective and unusually short in these cell lines. Likewise, this cell line initially had a translocation between chromosome 15 and chromosome 19 so it is likely that these chromosomes are also abnormal in these cell lines. The prevalence of fusion events involving chromosomes 19 and 22 may be due to the genetic profile of the HAP1 cell lines.

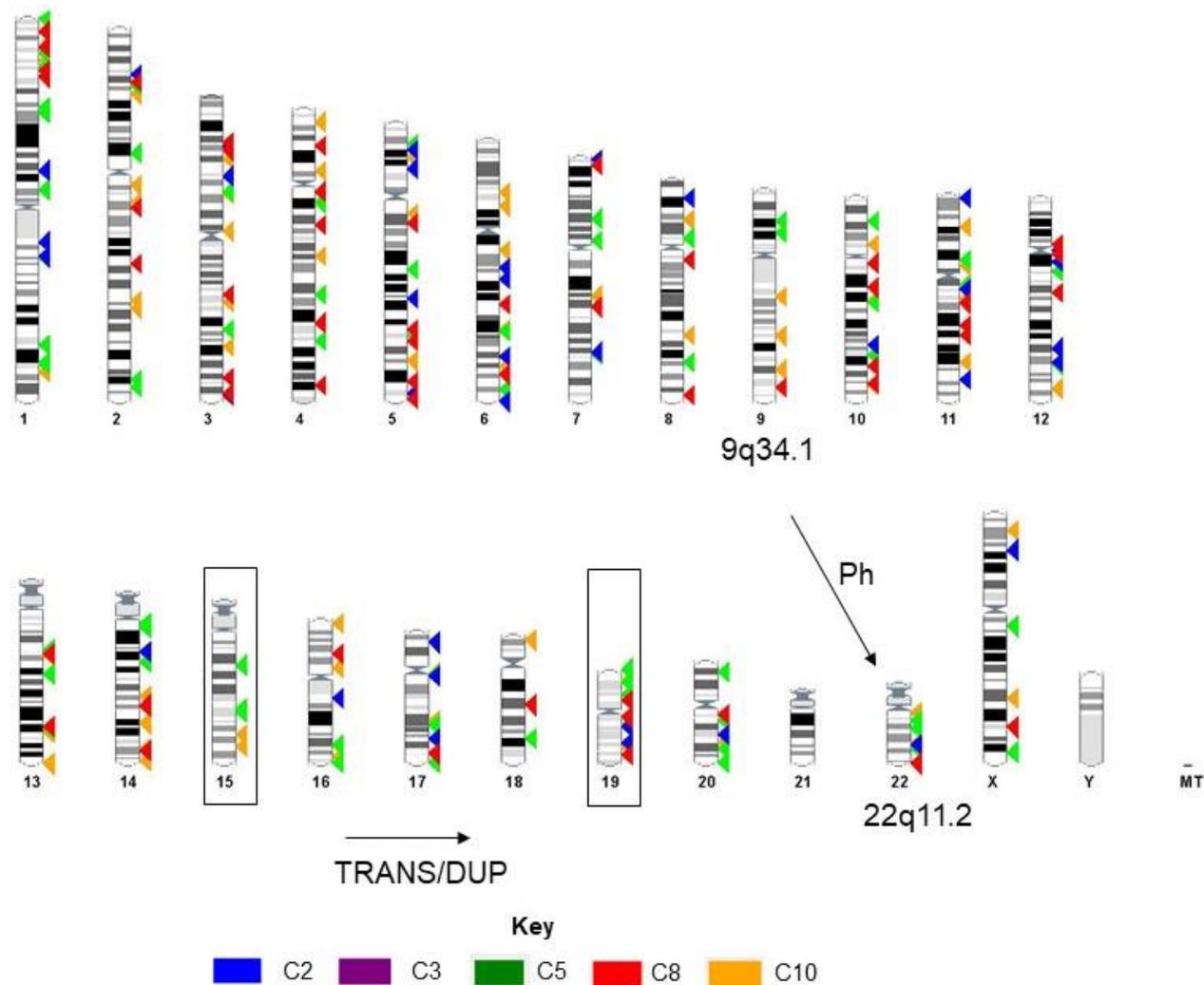
A



**Figure 5.5: (A)** Stacked bar charts showing the number of fusion events observed for HAP1 WT<sup>DN-hTERT</sup> (top) and HAP1 POLQ<sup>:DN-hTERT</sup> clones (bottom) following paired-end sequencing of fusion amplicons produced by 16p:21q multi-primer PCR.

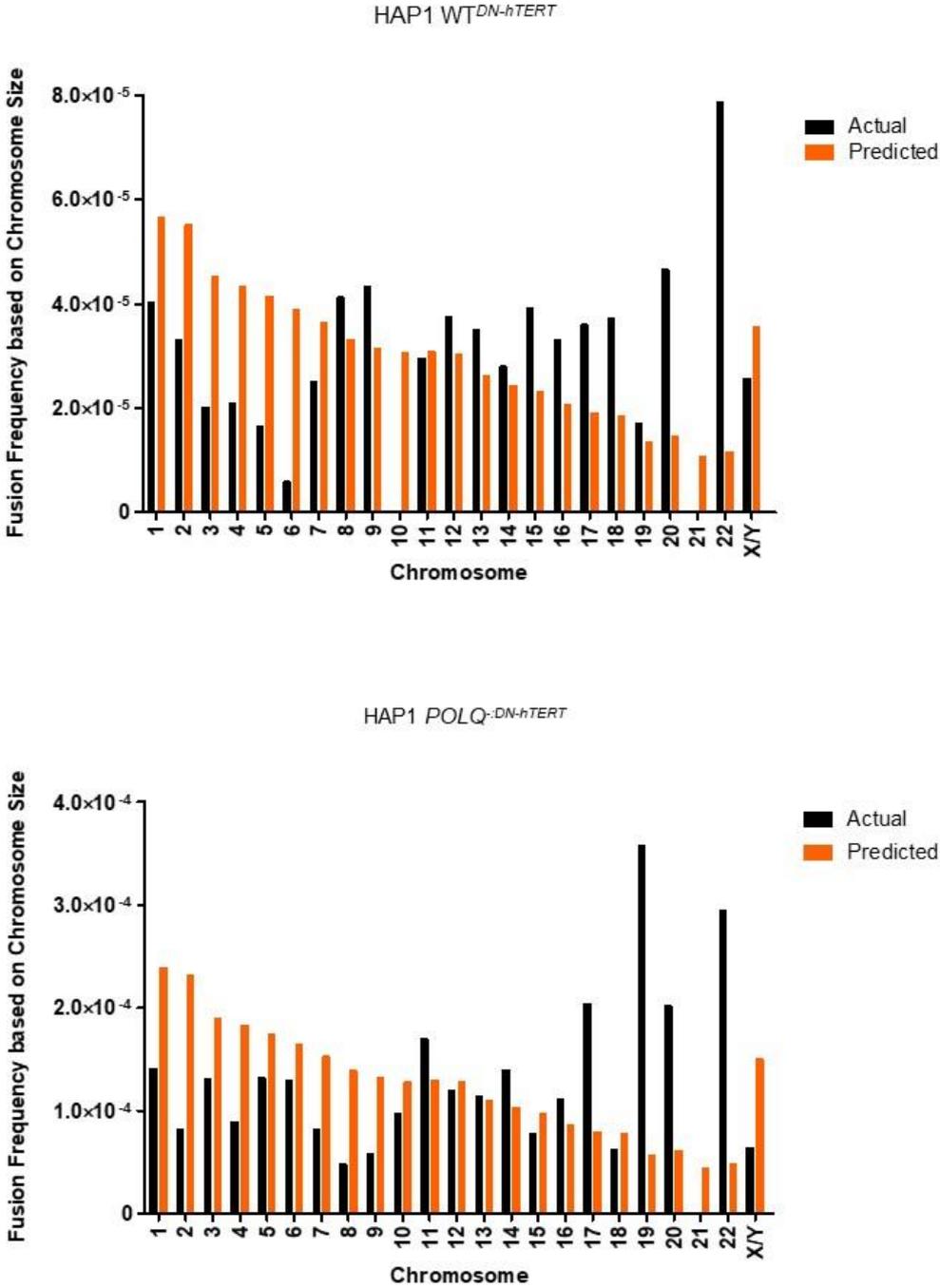
B





**Figure 5.5: (B)** Karyotype diagram (Ensembl) for validated genomic fusion events observed for HAP1 WT<sup>DN-hTERT</sup> (top) and HAP1 POLQ:<sup>DN-hTERT</sup> clones (bottom) following paired-end sequencing of fusion amplicons produced by 16p:21q multi-primer PCR. Each arrow represents a genomic fusion junction. Philadelphia translocation(Ph) locations highlighted in addition to existing duplication /translocation of chromosome 15/19 for these cell lines.

C



**Figure 5.5: (C)** Bar chart showing actual fusion frequency per chromosome of genomic telomere fusion events following paired-end sequencing compared with predicted fusion frequency per chromosome based on chromosome size for HAP1 WT<sup>DN-hTERT</sup> (top) and HAP1 POLQ<sup>-DN-hTERT</sup> cell lines (bottom).

### Fragile Sites

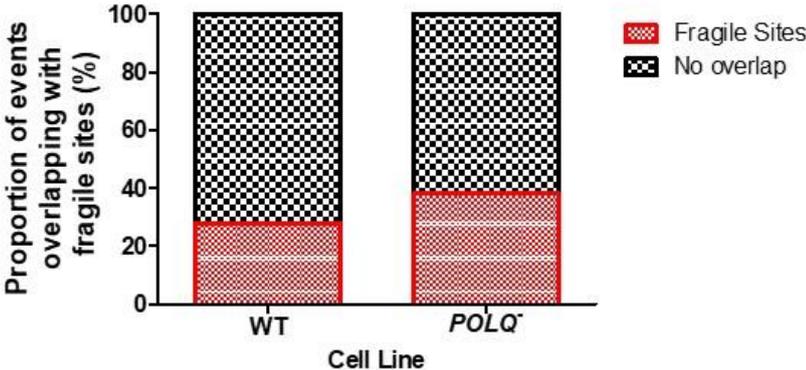
BED files (<https://webs.iiitd.edu.in/raghava/humcfs/index.html>) listing genomic locations of fragile sites were loaded into IGV to identify genomic fusion events that overlapped with fragile sites as read-pairs were visualised. For the HAP1 WT<sup>DN-hTERT</sup> cell line, 28% of these genomic fusions overlapped with fragile sites. Likewise, for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line, 38% of these genomic fusion events overlapped with fragile sites. There was therefore no significant difference in the proportion of genomic fusion events that overlap with fragile sites between the HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> cell lines ( $p=0.0805$ , Fisher's exact test) (**Figure 5.6A-B**).

### CpG Islands

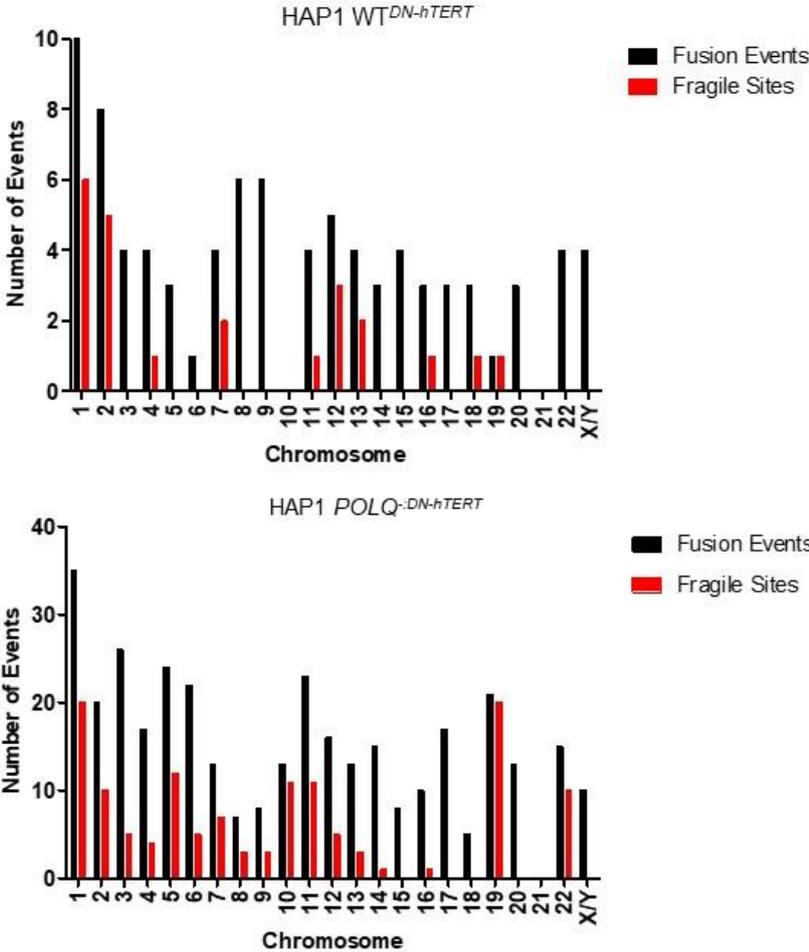
Any genomic fusion events that involved CpG islands were also documented to assess whether recognised chromatin states were linked to fusion frequency in a POLQ-deficient compared to a WT cell line.

No genomic fusion events involving CpG islands were observed for the HAP1 WT<sup>DN-hTERT</sup> cell line. For the HAP1 POLQ<sup>:DN-hTERT</sup> clones, a total of 8 genomic fusion events involving CpG islands were observed. This represents 2% of the total number of genomic fusion events recorded. Thus, when comparing the proportion of these events, there was no significant difference between the HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> cell lines ( $p=0.3659$ , Fisher's exact test). Therefore, genomic telomere fusion events were not associated with CpG islands.

A



B



**Figure 5.6:** (A) Stacked bar chart showing the proportion of genomic fusion events that overlap with fragile sites for HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>DN-hTERT</sup> clones. (B) Grouped bar charts showing the number of fusion events overlapping with fragile sites for HAP1 WT<sup>DN-hTERT</sup> (top) and HAP1 POLQ<sup>DN-hTERT</sup> clones (bottom).

## Genes

The involvement of protein-coding genes in the genomic fusion events identified was also recorded. For the HAP1 WT<sup>DN-hTERT</sup> clones, 55% of these genomic fusions involved genes compared to 61% of the genomic fusion events for the HAP1 POLQ<sup>-DN-hTERT</sup> clones (**Figure 5.7A**). This frequency is greater than expected based on the average gene content (43.4%) of the GRCh38.p13 reference human genome (*Chapter 2*). It was further documented whether these events were in exons or introns as previous data has demonstrated that inter-chromosomal fusion junctions were significantly more likely to occur within genes (Liddiard et al. 2016).

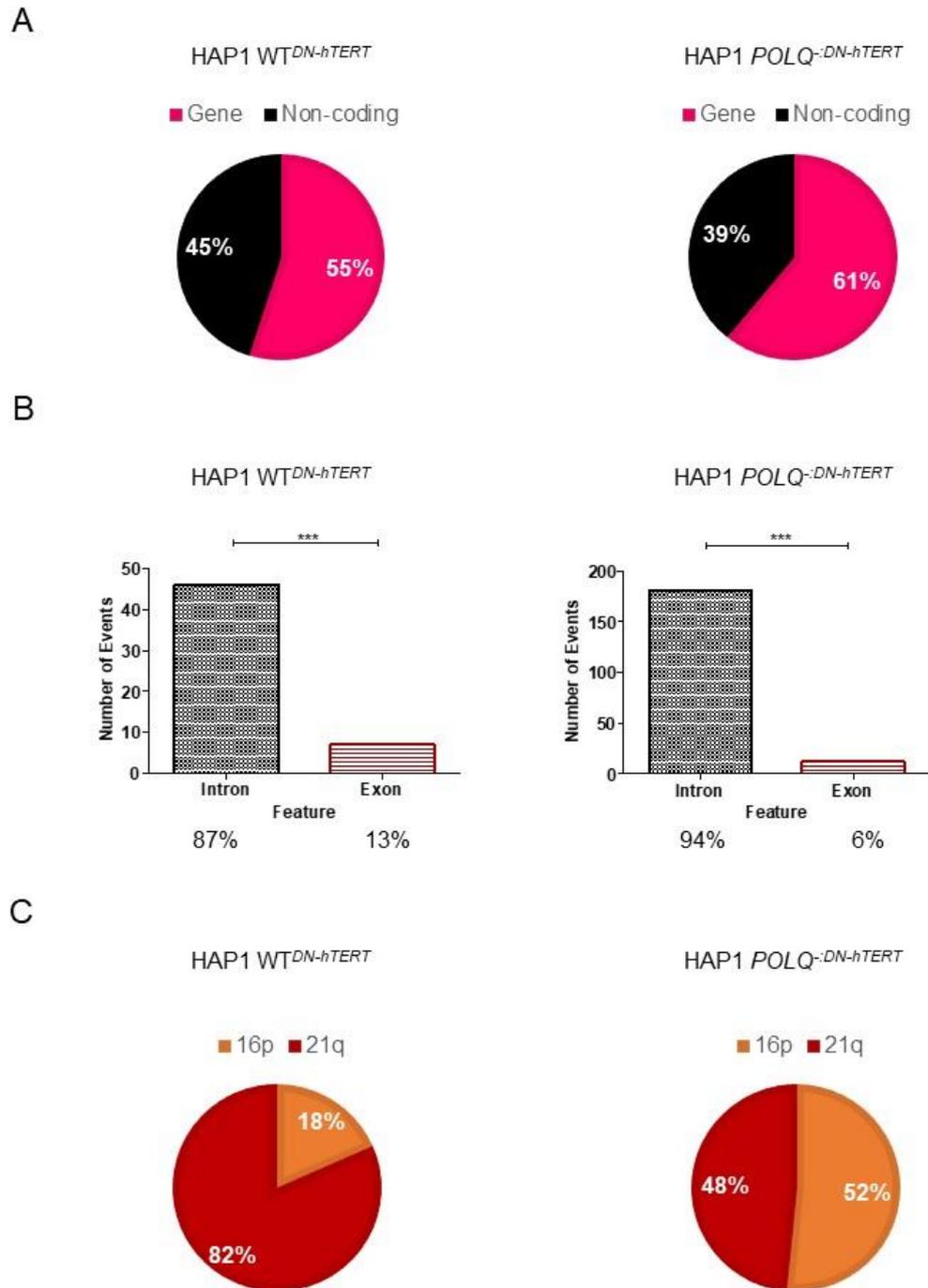
For the HAP1 WT<sup>DN-hTERT</sup> clones, a significantly greater abundance of the fusion junctions was in introns (87%) compared to in exons (13%,  $p \leq 0.001$ , Mann Whitney test). Similarly, a significantly greater abundance of these genomic fusion events that incorporated genes were in introns (94%), compared to exons (6%,  $p \leq 0.001$ , Mann Whitney test) for the HAP1 POLQ<sup>-DN-hTERT</sup> clones. There was no significant difference in the proportion of intron/exon involvement when comparing both cell lines ( $p=0.1407$ , Mann Whitney test) (**Figure 5.7B**). The greater abundance of events overlapping with introns is likely due to size as most introns are greater in length than exons (Sakharkar et al. 2004).

For the HAP1 WT<sup>DN-hTERT</sup> clones every genomic fusion event involving a gene was unique, that is, no gene identified was involved in more than one genomic fusion event. In contrast, as a larger number of genomic fusion events was recorded for the HAP1 POLQ<sup>-DN-hTERT</sup> clones, 15% of the total number of genes were involved in 2 separate genomic fusion events (*Appendix*).

## Subtelomeric family usage

Interestingly, and in conjunction with observations made in *Chapter 4*, 82% of the genomic fusion events for the HAP1 WT<sup>DN-hTERT</sup> clones were 21q linkages to the

genome compared to 18% of events which were 16p linkages to the genome. In contrast, 52% of the genomic fusion events observed for the HAP1 *POLQ*<sup>-DN-hTERT</sup> clones were 16p linkages to the genome whereas 48% involved the 21q linkages to the genome (**Figure 5.7C**). This reflects what was observed for the Southern blots generated following 16p:21q fusion PCR for these clones (*Chapter 4, Section 4.4.3, Figure 4.4A*).



**Figure 5.7:** (A) Pie charts showing proportion of genomic fusion junctions that overlapped with genes for HAP1 WT<sup>DN-hTERT</sup> (left) and HAP1 POLQ<sup>-DN-hTERT</sup> clones (right). (B) Bar chart showing proportion of genomic fusion events involving genes that are in introns or exons for HAP1 WT<sup>DN-hTERT</sup> (left) and HAP1 POLQ<sup>-DN-hTERT</sup> clones (right). Significance was determined using a Fisher's exact test. (C) Pie charts showing proportion of genomic fusion events involving 16p or 21q telomere families for HAP1 WT<sup>DN-hTERT</sup> (left) and HAP1 POLQ<sup>-DN-hTERT</sup> clones (right).

## Gene Function

Gene function analysis was performed using the PANTHER classification system (Mi et al. 2019) to determine whether there were any commonalities amongst the genes involved in the telomere fusion events recorded.

## Molecular Function

For the HAP1 WT<sup>DN-hTERT</sup> cell line, the most common molecular functions associated with the genes involved in the fusion events were catalytic activity (34.9%) and binding (30.2%). These are presented as a percentage of the total number of genes recorded. Similarly, for the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line, the most common molecular functions were binding (34%) and catalytic activity (24.1%) (**Figure 5.8A**).

## Biological Processes

The most common biological process associated with the genes involved in the telomere fusion events documented for both the HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line was 'cellular processes' (37.2% and 36.2% respectively). These include processes such as response to stimuli, signal transduction and cellular organisation. Likewise, metabolic processes (23.3% and 24.1%) and biological regulation (16.3% and 24.1%) were amongst the most common categories for both cell lines (**Figure 5.8B**).

## Pathway

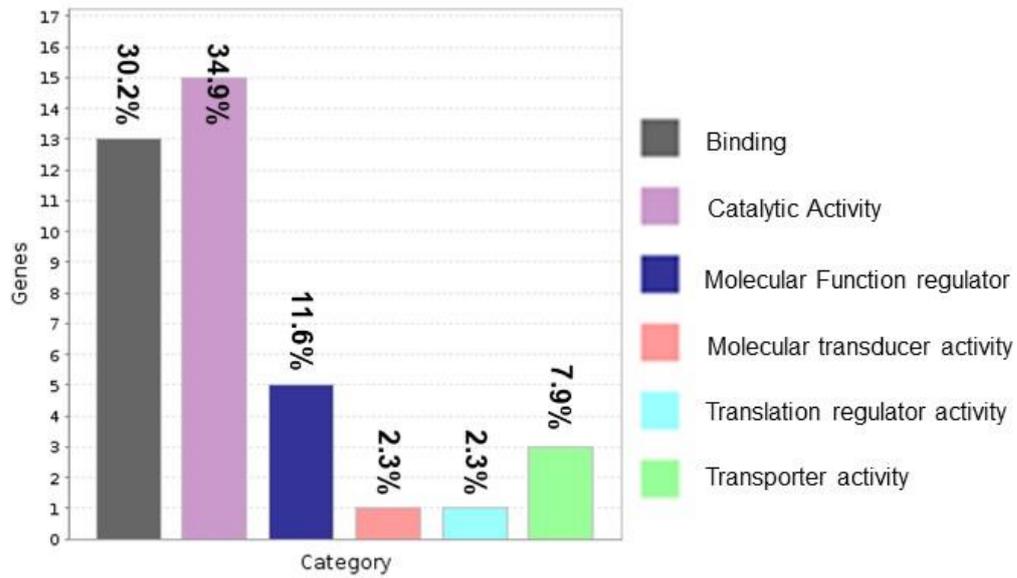
Due to the greater number of genomic telomere fusion events observed for the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line compared with the HAP1 WT<sup>DN-hTERT</sup> cell line, a larger number of pathways (9 and 40 respectively) associated with the genes involved in telomere fusion events were also recorded. Of the 9 pathways identified for the HAP1 WT<sup>DN-hTERT</sup> cell line, 8 of them represented 2.3% of the total number of genes recorded. However, genes associated with cholecystokinin receptor (CCKR) signalling represented 4.7% of the total number of genes recorded. Of the 9 signalling pathways identified for the HAP1 WT<sup>DN-hTERT</sup> cell line, 6 were also identified for the HAP1 *POLQ*<sup>-DN-hTERT</sup>

*:DN-hTERT* cell line. The most common pathway identified for the HAP1 *POLQ<sup>:DN-hTERT</sup>* cell line was the Wnt signalling pathway. Genes associated with this pathway accounted for 5% of the total number of genes recorded for all genomic telomere fusion events. This pathway was also associated with genes identified for the HAP1 *WT<sup>DN-hTERT</sup>* cell line. Similarly, the next most common pathway identified for the HAP1 *POLQ<sup>:DN-hTERT</sup>* cell line was the CCKR signalling pathway which represented 3.5% of the total number of genes recorded. (*Appendix*).

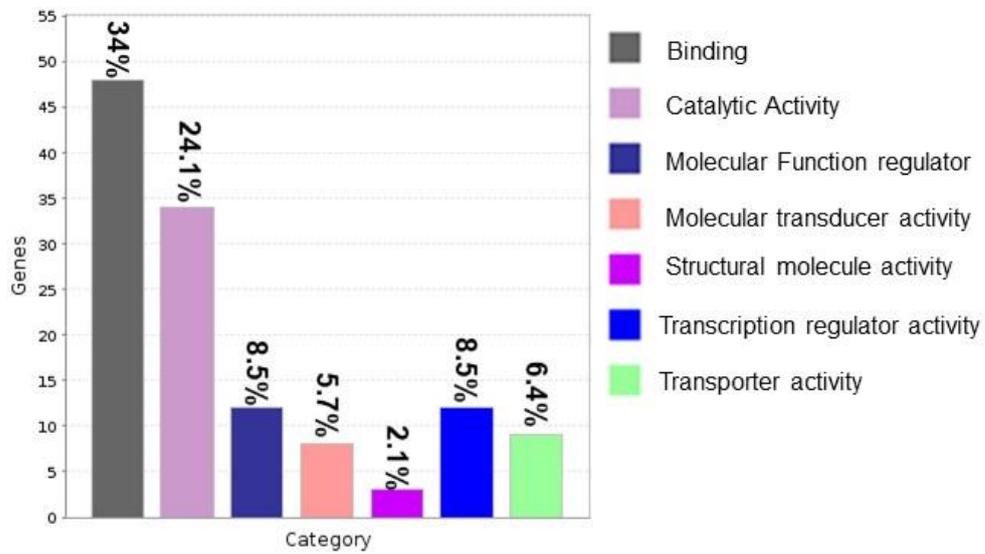
A

Molecular Function

HAP1 WT<sup>DN-hTERT</sup>

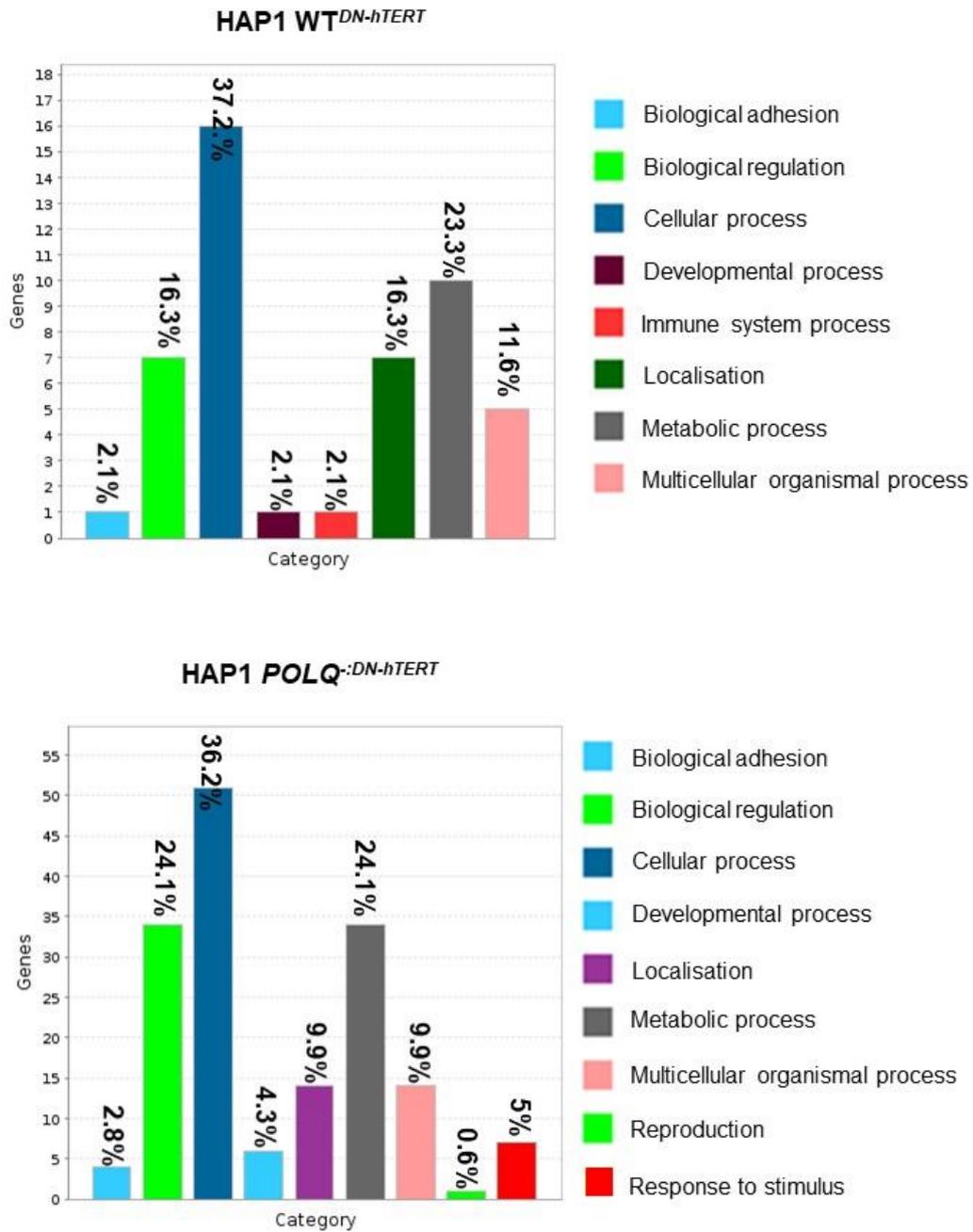


HAP1 POLQ<sup>DN-hTERT</sup>



B

## Biological processes



**Figure 5.8: (A)** Bar charts showing the proportion of genes overlapping with telomere fusion junctions that are associated with different molecular functions and biological processes **(B)** for HAP1 WT<sup>DN-hTERT</sup> (top) and HAP1 POLQ<sup>DN-hTERT</sup> cell lines (bottom)

#### 5.4.8 16p – 21q fusion events

The 16p-21q family fusion events represent fusion between telomere/subtelomeric sequences not involving non-telomeric loci. 16p-21q family fusion event reads were visualised using IGV and BLAST-alignment was used to manually validate the events and determine the fusion junction location. Fusion junction features including microhomology and insertions were also recorded and manually validated.

As stated previously, the first 100 events per clone were manually validated (based on likelihood of true event, determined by machine learning program). However, only 18 inter-chromosomal 16p-21q events were recorded for all HAP1 WT<sup>DN-hTERT</sup> clones (between 1 and 8 events were recorded for each clone). Following validation, the total number of events for the HAP1 WT<sup>DN-hTERT</sup> clones was only 11 (between 1 and 4 events per clone). In contrast, 842 events were recorded for the HAP1 POLQ<sup>-DN-hTERT</sup> clones, ranging from 133 to 232 events detected per clone. Following validation of the first 100 events per clone, a total of 395 events were recorded (between 70 and 86 events per clone). Thus, a significantly greater number of 16p-21q family fusion events was observed for the HAP1 POLQ<sup>-DN-hTERT</sup> clones compared to the HAP1 WT<sup>DN-hTERT</sup> clones ( $p < 0.05$ , Mann Whitney test) (**Figure 5.9A**). This reflects previous data obtained following 16p:21q fusion PCR which demonstrated a significantly greater frequency of 16p-21q inter-chromosomal fusion events for the HAP1 POLQ<sup>-DN-hTERT</sup> compared with the HAP1 WT<sup>DN-hTERT</sup> cell line (*Chapter 4, Section 4.4.3, Figure 4.3*).

#### Fusion Junction Features

Fusion junction sequences were analysed to determine whether there was any difference in microhomology usage or the abundance of insertions, although only 11 16p-21q valid events were recorded in total for the HAP1 WT<sup>DN-hTERT</sup> clones. Fusion junctions with no microhomology (blunt fusion junctions) were the most common fusion junction feature (45.5%, 5 events) compared with fusion junctions containing microhomology (36.4%, 4 events) and fusion junctions with insertions (18.2%, 2

events). There was therefore no significant difference in the proportions of the different junction features for this cell line ( $p=0.6351$  for MH and B compared to INS,  $p=1.00$  for MH compared to B, Fisher's exact test) (**Figure 5.9B**)

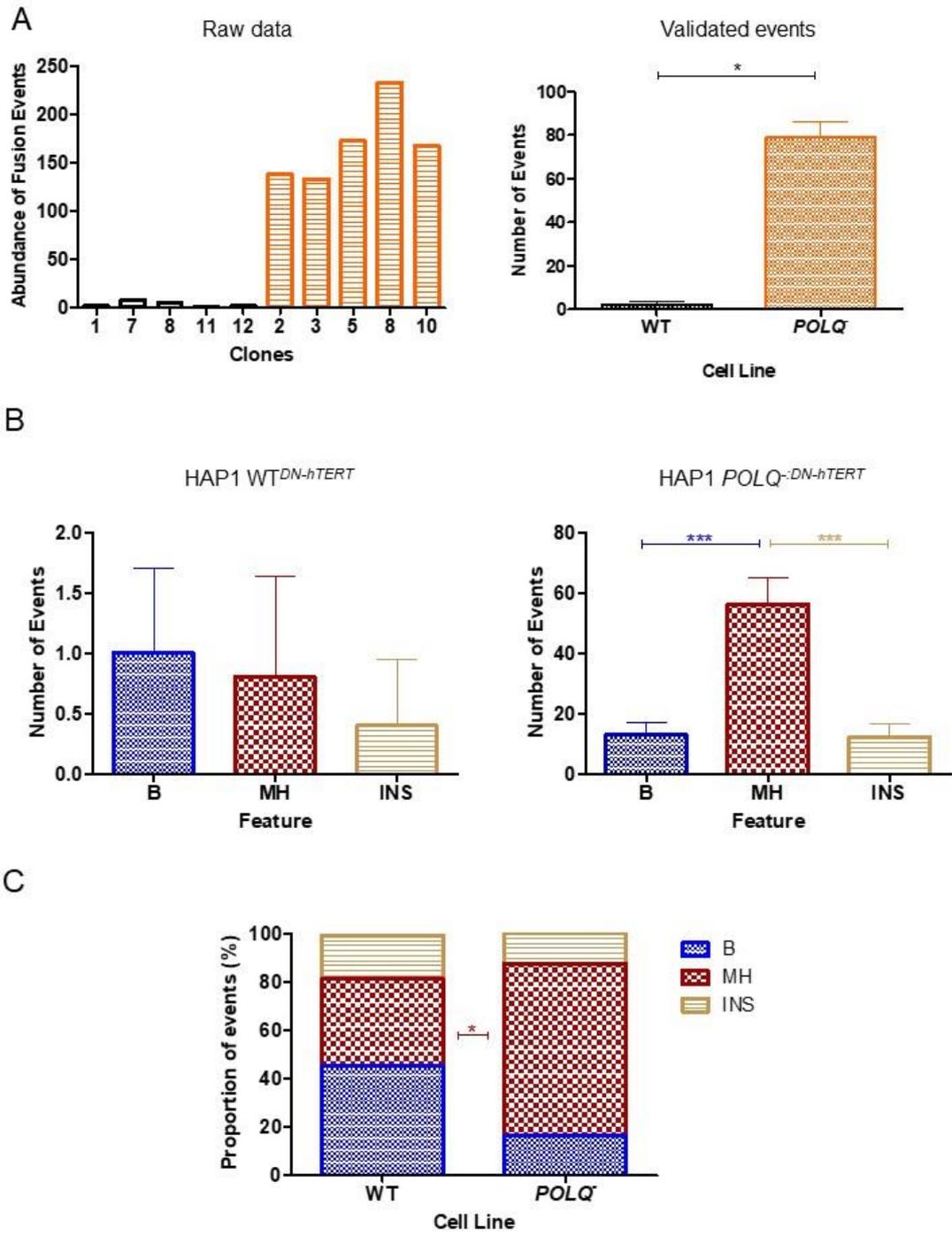
For the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones, fusion junctions with microhomology were the most frequently observed (71.1%, 281 events) compared with blunt fusion junctions (16.5%, 65 events) and fusion junctions with insertions (12.4%, 49 events). As such, a significantly greater proportion of fusion junctions with microhomology were recorded compared with both blunt fusion junctions and fusion junctions with insertions ( $p<0.001$ , Fisher's exact test) (**Figure 5.9C**). No significant differences were observed when comparing the proportion of 16p-21q and genomic fusion junctions with microhomology (71% compared with 57.5% for the genomic events) or insertions (16% versus 9.7% for genomic events). However, a 2.73-fold decrease in the proportion of blunt fusion junctions (12% compared with 32.8% for genomic events) was observed when comparing 16p-21q and genomic fusion events for this cell line.

A significantly greater proportion of fusion junctions with microhomology was observed for the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones compared to the HAP1 *WT*<sup>DN-hTERT</sup> clones ( $p<0.05$ , Fisher's exact test). The mean length of the microhomology for the HAP1 *WT*<sup>DN-hTERT</sup> clones was 3.3 bp compared to mean length of 2.3 bp for the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones, thus there was no significant difference in microhomology length for the two different cell lines ( $p=0.3033$ , Mann Whitney test).

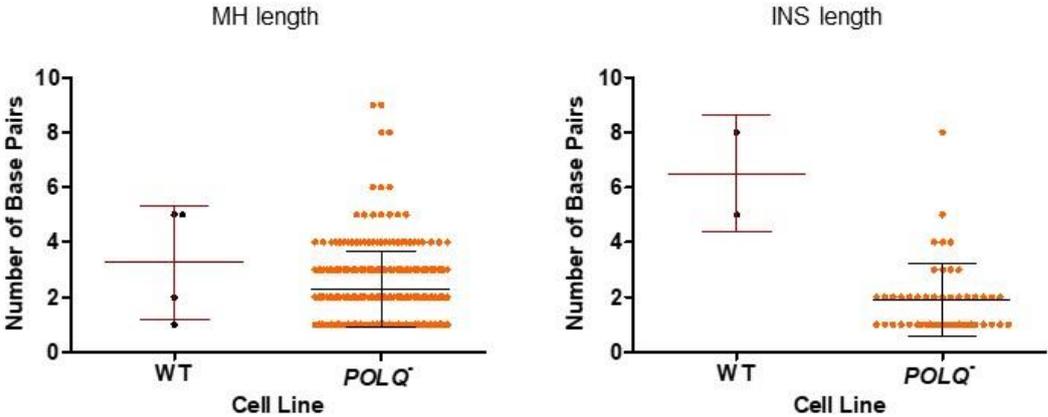
Despite no significant difference in the proportion of fusion junctions with insertions between the HAP1 *WT*<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>:DN-hTERT</sup> clones, the length of the insertions varied. The mean length of the insertions for the HAP1 *WT*<sup>DN-hTERT</sup> clones was 6.5 bp compared to 2.2 bp for the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones (**Figure 5.9D**). This reflects observations made for genomic fusion events described previously.

Insertion sequences were categorised as templated insertions if the exact insertion sequence was mapped to within 100 bp of the fusion junction. Insertions  $\leq 2$  bp were categorised as 'undetermined'. Insertions  $\geq 3$  bp were analysed to determine whether they were templated or non-templated. For the HAP1 WT<sup>DN-hTERT</sup> cell line, only 2 fusion junctions with insertions were detected, 1 of these insertions was templated (50%) whereas the other event was non-templated (50%). For the HAP1 POLQ<sup>:DN-hTERT</sup> cell line, 49 fusion junctions with insertions were identified. The greatest proportion of events were  $\leq 2$  bp and therefore categorised as undetermined (40 events, 82%). 6 fusion junctions (12%) had templated insertions compared to 3 fusion junctions with non-templated insertions (6%).

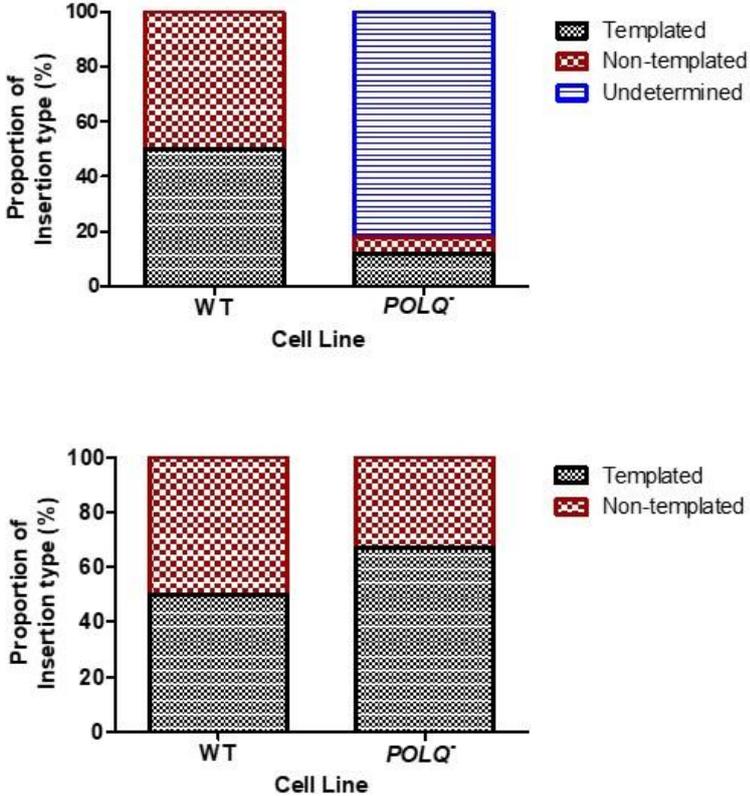
When comparing the proportion of templated and non-templated insertions only for the two cell lines, there were no significant differences ( $p=1.00$ , Fisher's exact tests) in the proportion of templated (50% HAP1 WT<sup>DN-hTERT</sup>, 67% HAP1 POLQ<sup>:DN-hTERT</sup>) or non-templated events (50% HAP1 WT<sup>DN-hTERT</sup>, 33% HAP1 POLQ<sup>:DN-hTERT</sup>) (**Figure 5.9E**).



D



E



**Figure 5.9: (A)** Bar chart showing number of 16p-21q fusion events before validation and mean number of manually validated 16p-21q fusion events for HAP1<sup>DN-hTERT</sup> cell lines. Error bars indicate standard deviation and significance determined using a Mann Whitney test. **(B)** Bar charts showing the mean number of fusion junction features for HAP1<sup>DN-hTERT</sup> cell lines. Error bars indicate standard deviation and significance was determined using a Mann Whitney test. **(C)** Stacked bar chart showing the proportion of fusion junction features for HAP1<sup>DN-hTERT</sup> cell lines. Significance was determined using a Fisher's exact test **(D)** Scatter plot showing the length of microhomology (left) and insertions (right) at 16p-21q fusion junctions for HAP1 WT<sup>DN-hTERT</sup> and POLQ<sup>-</sup>:DN-hTERT clones. Error bars indicate the mean and standard deviation values. **(E)** Stacked bar charts showing the proportion of undetermined (top), templated and non-templated (bottom) insertions at fusion junctions in HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-</sup>:DN-hTERT cell lines. Significance was determined using a Fisher's exact test.

#### 5.4.9 '16p' and '21q' events

As there is a redundancy in 16p and 21q primer targets, it is not possible to discern whether these '16p' / '21q' events are intra- or inter-chromosomal fusion events due to the shared sequence similarity within these telomere families.

As with the 16p-21q inter-chromosomal fusion events, the first 100 (based on likelihood of true event, determined by machine learning program) '16p' / '21q' fusion events were manually validated for each clone. Following this validation, the total number of '16p' / '21q' events recorded for the HAP1 WT<sup>DN-hTERT</sup> clones was 382, ranging from 70-85 events per clone. A total of 453 events was detected for the HAP1 POLQ<sup>:DN-hTERT</sup> clones, with the numbers of events per clone ranging from 87-95. As with the genomic and inter-chromosomal fusion events, significantly more '16p' / '21q' events were observed for the HAP1 POLQ<sup>:DN-hTERT</sup> clones compared with the HAP1 WT<sup>DN-hTERT</sup> clones ( $p < 0.01$ , Mann Whitney test) (**Figure 5.10A**).

Interestingly, '21q' events accounted for 25% of these events, although there were 3 x more '16p' events which represented 75% of these events. Similarly, for the HAP1 POLQ<sup>:DN-hTERT</sup> clones, '16p' events also made up 92% of these events, compared with 8% for '21q' events, this equates to an 11.5-fold greater abundance of 16p only events compared with '21q' events (**Figure 5.10B**). In contrast, 21q-linked genomic events represented 82% and 48% compared with 18% and 52% 16p-linked genomic events for the HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> cell lines respectively. (Section 5.4.7).

#### Fusion junction Features

For the HAP1 WT<sup>DN-hTERT</sup> clones, fusion junctions with microhomology were the most common (89%) and represented a significantly greater proportion of events compared with both blunt fusion junctions (5%) and fusion junctions with insertions (6%) ( $p < 0.001$ , Fisher's exact tests). Likewise, fusion junctions with microhomology were also the most common for the HAP1 POLQ<sup>:DN-hTERT</sup> clones (95%) and represented a significantly greater proportion of events compared with both blunt fusion junctions

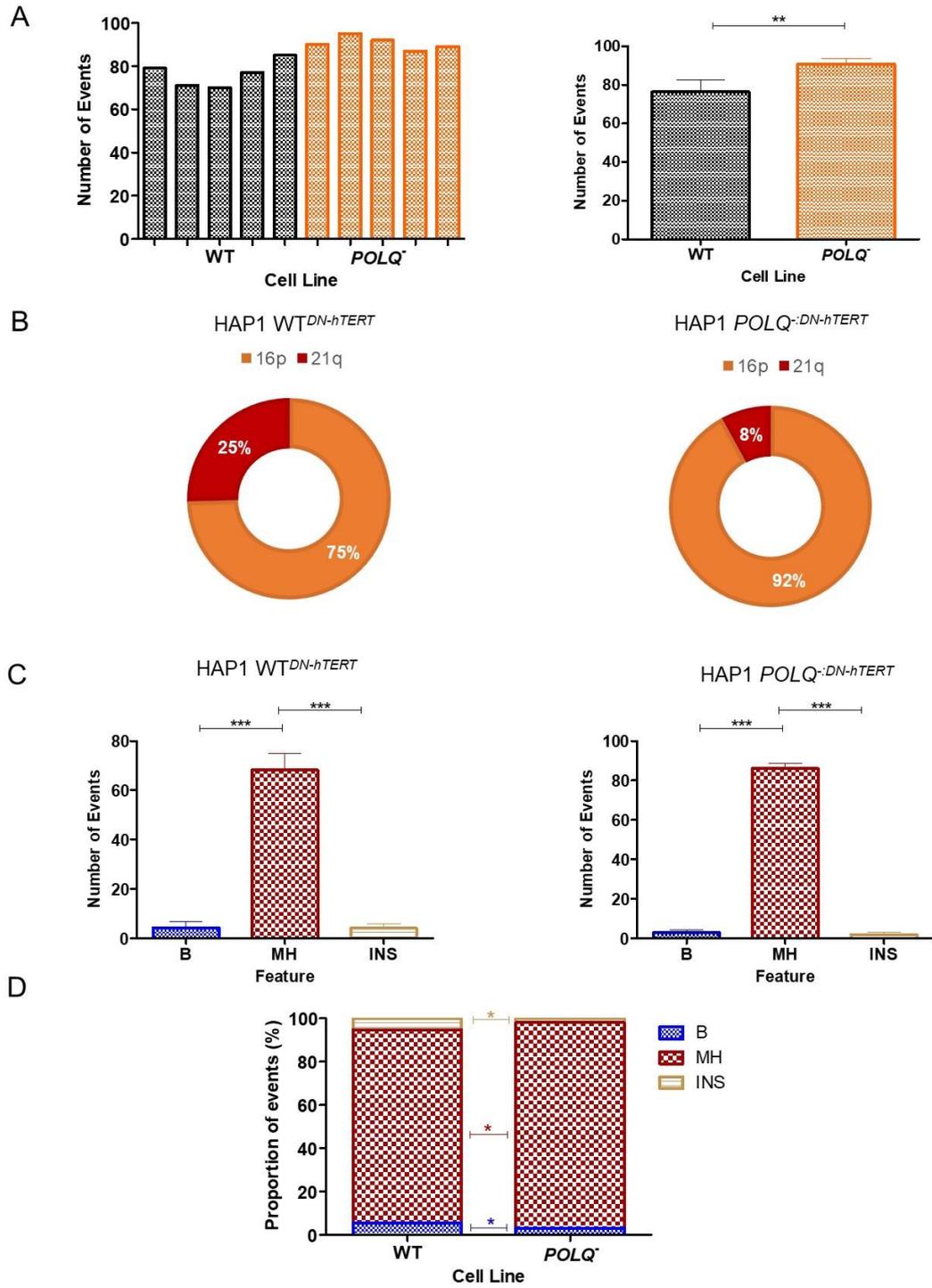
(2%) and fusion junctions with insertions (3%) ( $p < 0.001$ , Fisher's exact tests). (**Figure 5.10C**). An increase in the proportion of fusion events with microhomology at the fusion junction was observed compared with genomic and 16p-21q events for both cell lines.

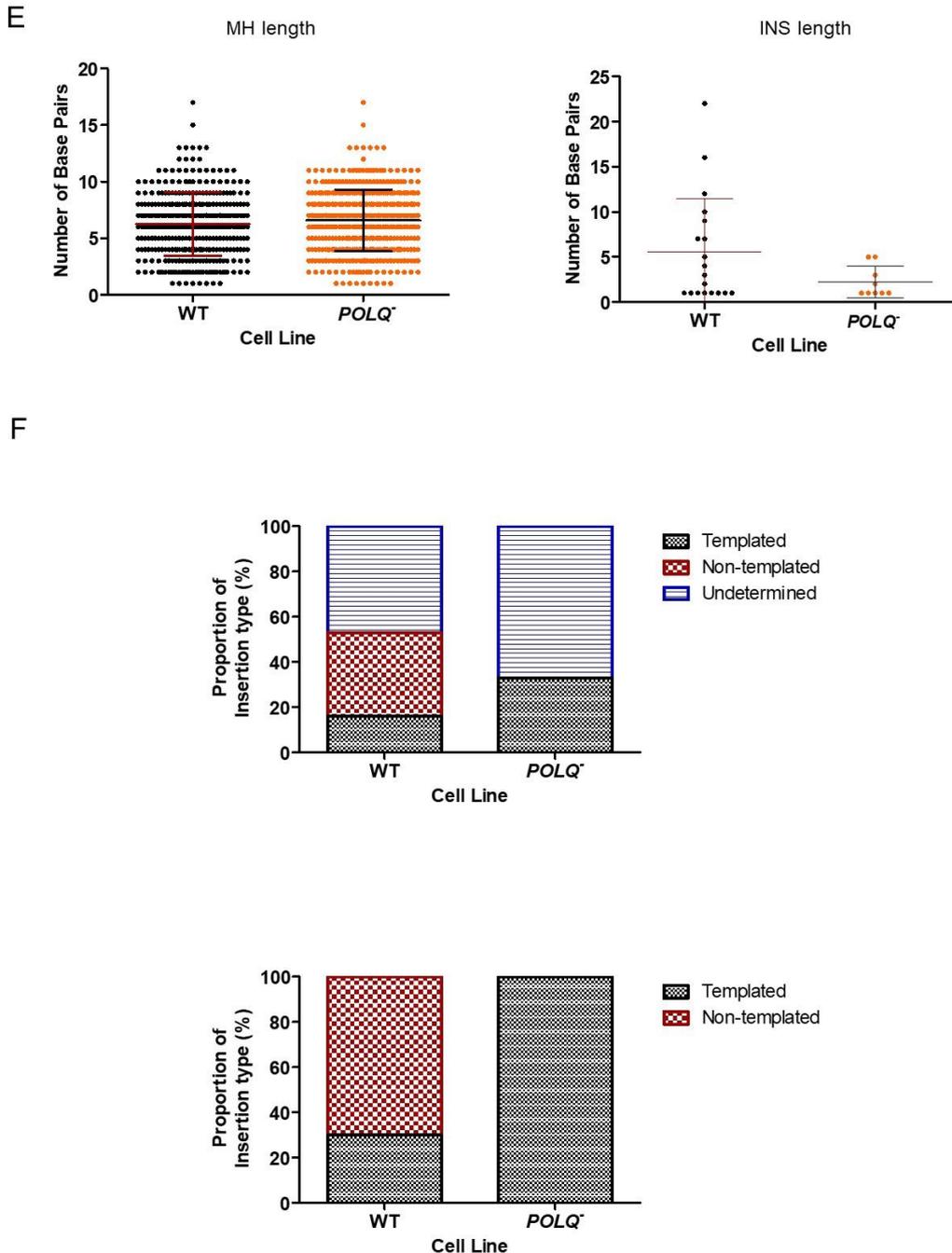
Although junctions with microhomology were most common for both the HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> cell lines, the proportion of these junctions was significantly greater for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line ( $p \leq 0.001$ , Fisher's exact test). As such, the proportion of blunt junctions and junctions with insertions was significantly greater in the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line ( $p \leq 0.05$ , Fisher's exact test) (**Figure 5.10D**).

However, when comparing the length of the insertion or the microhomology observed, no significant differences were seen between the two cell lines ( $p = 0.0763$  and  $p = 0.3054$  respectively, Mann Whitney test). A greater range of insertion lengths were recorded for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line, reflected by the standard deviation values for the distributions of 5.83 and 1.62 bp respectively (**Figure 5.10E**).

Insertions at the fusion junction were categorised as templated, non-templated or undetermined as previously described. For the HAP1 WT<sup>DN-hTERT</sup> cell line, 47% of junctions with insertions were undetermined (9 events) compared with 37% of events with templated insertions (7 events) and 16% non-templated insertions (3 events). When analysing just templated and non-templated insertions, 70% were non-templated whereas 30% were templated insertions.

For the HAP1 POLQ<sup>:DN-hTERT</sup> cell line, 67% of insertion events were undetermined (6 events) whereas 33% were templated insertions (3 events). As such, templated insertions accounted for 100% of the insertions when analysing templated with non-templated insertions. However, there was no significant difference when comparing the proportion of templated or non-templated insertions between the two cell lines ( $p = 0.699$ , Fisher's exact tests) (**Figure 5.10F**).





**Figure 5.10:** (A) Chart showing the total number of ‘16p’ or ‘21q’ events recorded from raw data (left) and the mean number of manually validated events (right) for HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>-DN-hTERT</sup> cell lines. Error bars indicate standard deviation and significance was determined using a Mann Whitney test. (B) Pie charts showing the proportion of ‘16p’ compared with ‘21q’ events for HAP1 WT<sup>DN-hTERT</sup> (left) and HAP1 *POLQ*<sup>-DN-hTERT</sup> (right) cell lines. (C) Charts showing the mean numbers of blunt (B) fusion junctions, junctions with microhomology (MH) and junctions with insertions (INS) for HAP1 WT<sup>DN-hTERT</sup> clones and HAP1 *POLQ*<sup>-DN-hTERT</sup> clones. Error bars indicate standard deviation and significance was determined using a Mann Whitney test. (D) Stacked bar chart showing the proportion of fusion junction features for the two cell lines. (E) Scatter plot showing length of microhomology (left) and insertions (right) for the HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>-DN-hTERT</sup> cell lines. Error bars indicate the mean and standard deviation values and significance was determined using a Mann Whitney test. (F) Stacked bar chart showing the proportion of junction types for the HAP1 WT<sup>DN-hTERT</sup> compared with the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line. Significance was determined using a Fisher’s exact test.

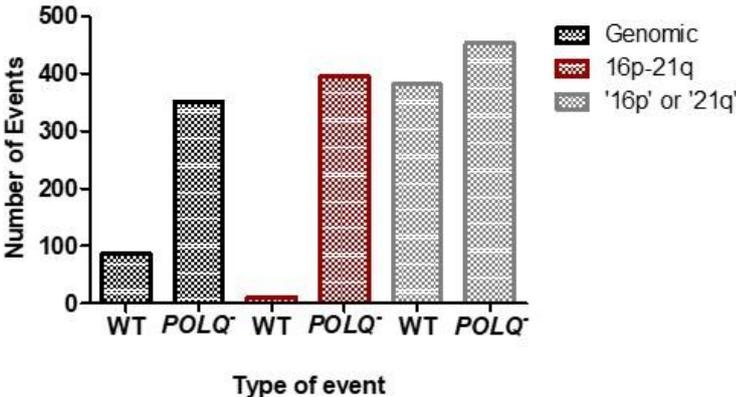
#### 5.4.10 Overall fusion profiles

For the HAP1 WT<sup>DN-hTERT</sup> cell line, the greatest proportion of fusion events observed was '16p' or '21q' fusion events (80%, 382 events). Therefore, these events represent a significantly greater proportion of events compared with both genomic (18%, 87 events) and 16p-21q inter-chromosomal fusion events (2%, 11 events) observed ( $p \leq 0.001$ , Fisher's exact test). The difference in the proportion of genomic fusion events compared to 16p-21q inter-chromosomal events was also significant ( $p \leq 0.001$ , Fisher's exact test).

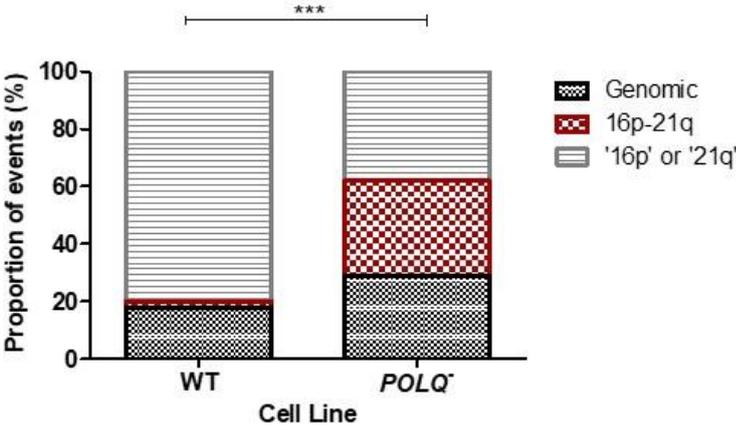
Similarly, '16p' or '21q' fusion events also represented the greatest proportion of fusion events for the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line (38%, 453 events), although there was no significant difference in the proportion of these events compared with 16p-21q inter-chromosomal fusion events (33%, 395 events) ( $p = 0.056$ , Chi-squared test). However, both event types accounted for significantly greater proportion of events compared with genomic fusion events (29%, 351 events) ( $p \leq 0.001$ ,  $p < 0.05$ , respectively, Chi-squared test) (**Figure 5.11A**).

Whilst '16p' or '21q' fusion events represented the greatest proportion of fusion events for both cell lines, the proportion of these events in the HAP1 WT<sup>DN-hTERT</sup> cell line was significantly greater when compared to the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line ( $p < 0.001$ , Fisher's exact test). As these '16p' or '21q' fusion events represented such a large proportion of events in the HAP1 WT<sup>DN-hTERT</sup> cell line, the proportion of both genomic and 16p-21q inter-chromosomal events was significantly greater in the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line ( $p < 0.001$ , Fisher's exact test) (**Figure 5.11B**).

A



B



**Figure 5.11: (A)** Bar chart showing the number of validated genomic, 16p-21q and '16p' / '21q' fusion events recorded for HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-DN-hTERT</sup> cell lines. **(B)** Stacked bar chart comparing proportions of genomic, 16p-21q and '16p' or '21q' fusion events for HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-DN-hTERT</sup> cell lines. Significance was determined using a Fisher's exact test (comparing each type of event between cell lines).

#### 5.4.11 Whole Genome Sequencing Analysis

To determine the potential contribution of telomere fusions to the global genomic instability occurring in crisis, 8 HAP1 WT<sup>DN-hTERT</sup> and 8 HAP1 POLQ<sup>-DN-hTERT</sup> clones were sampled before and after crisis for whole genome paired-end sequencing in addition to parental lines. The mapping and initial analyses were performed by Dr Kez Cleal, Cardiff University. The 5 clones for each cell line that were subjected to fusion amplicon sequencing were also used for this analysis, in addition to a further 3 clones for each cell line. To ensure samples were taken before the onset of crisis, samples between passage 5-8 were selected for each clone, these time points were associated with continued telomere erosion before the onset of crisis, *Chapter 3*. Likewise, to ensure samples were taken following escape from crisis, samples between passage 32 and 48 were chosen for each clone, these time points were associated with telomerase-mediated telomere length stabilisation, *Chapter 3*. It should be noted that telomere fusion analyses previously performed were using samples that were transiting telomere-driven crisis and thus represent transient fusion events that cannot be detected following the reactivation of telomerase and the escape from crisis.

#### Clonality

A difference in the growth curves of the clones going through crisis was observed, with POLQ-deficient clones transiting crisis more rapidly than the WT cell lines (most obvious in HCT116) (*Chapter 3*). This observation is consistent with more single cell clones being capable of escaping crisis in the absence of POLQ. Therefore, these clonal populations were analysed to determine the clonality of the sample before and after crisis. A monoclonal culture is defined as a population without significant SVs or allelic divergences, therefore, all the cells in a monoclonal culture are uniform in terms of variants. Polyclonal cultures are defined by divergences either in large scale SVs or SNPs/SNVs. A difference in clonality may have been caused by increased variability amongst cell populations, possibly due to an increased variety of telomere

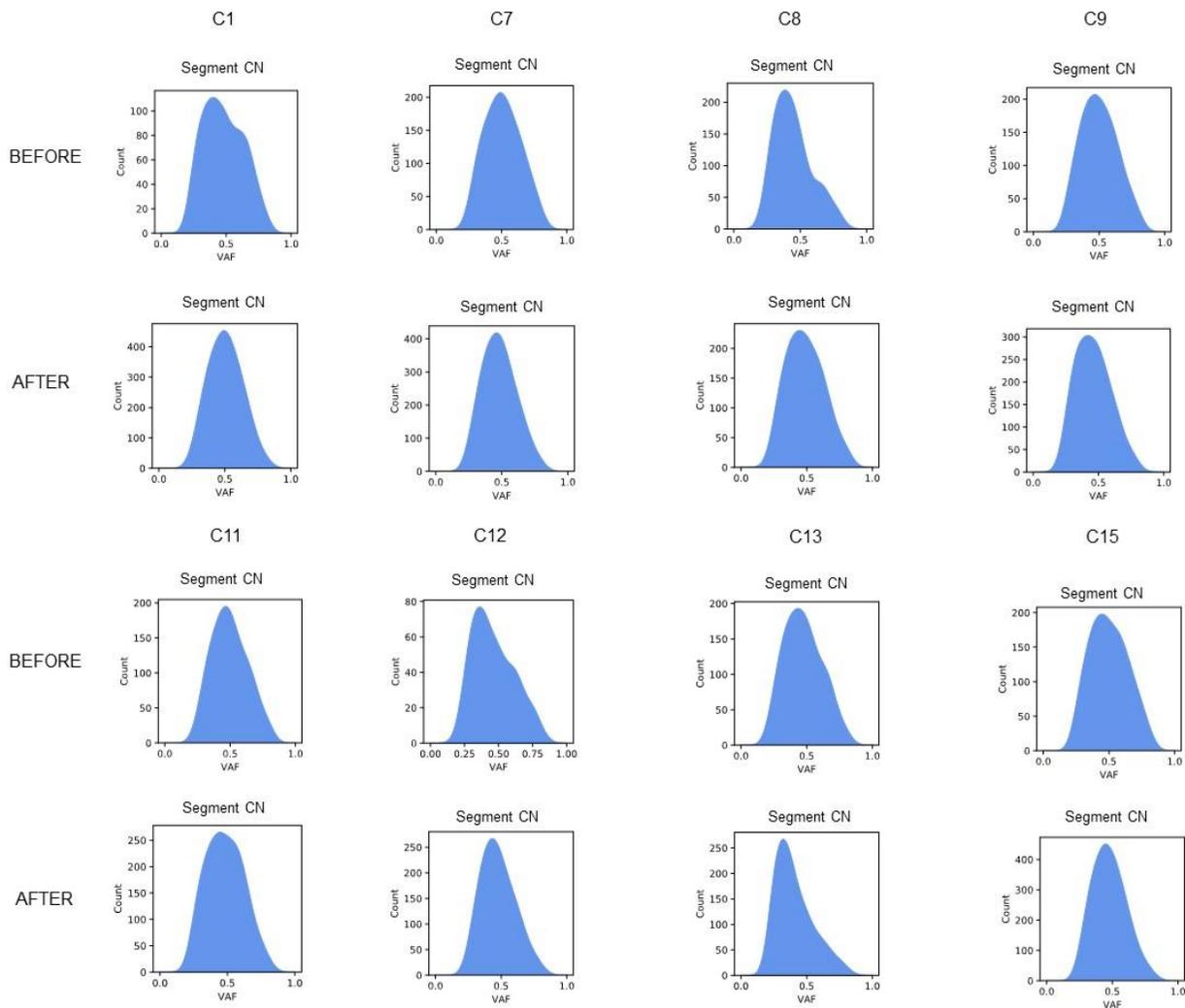
fusion events, which could facilitate the reactivation of telomerase and the escape from crisis. As such, unique variant allele frequency (VAF) profiles in diploid regions were assessed (Cleal et al. 2019) and using principles outlined below, the clonality of cultures was manually inferred. For a monoclonal culture, the VAF is 0.5 whereas additional subpopulations associated with a polyclonal culture will cause a left-shift in the VAF peak. For a culture with two equally represented subclones, a novel heterozygous variant in this population would have an allele fraction of 0.25, thus shifting the VAF peak towards lower values. Therefore, a lower VAF is representative of polyclonality of the culture (**Figure 5.12A**). As the ploidy of these samples is not fully understood, any left-shift in the VAF was interpreted as representing a polyclonal culture.

Highlighting specific examples, the HAP1 *POLQ*<sup>:DN-hTERT</sup> C8 after crisis sample was considered to be polyclonal as many unique single nucleotide variants (SNVs) were identified resulting in a robust left-shifted VAF profile. The HAP1 WT<sup>DN-hTERT</sup> C1 before and after crisis samples were deemed to be monoclonal as a VAF peak at 0.5 was identified for both samples.

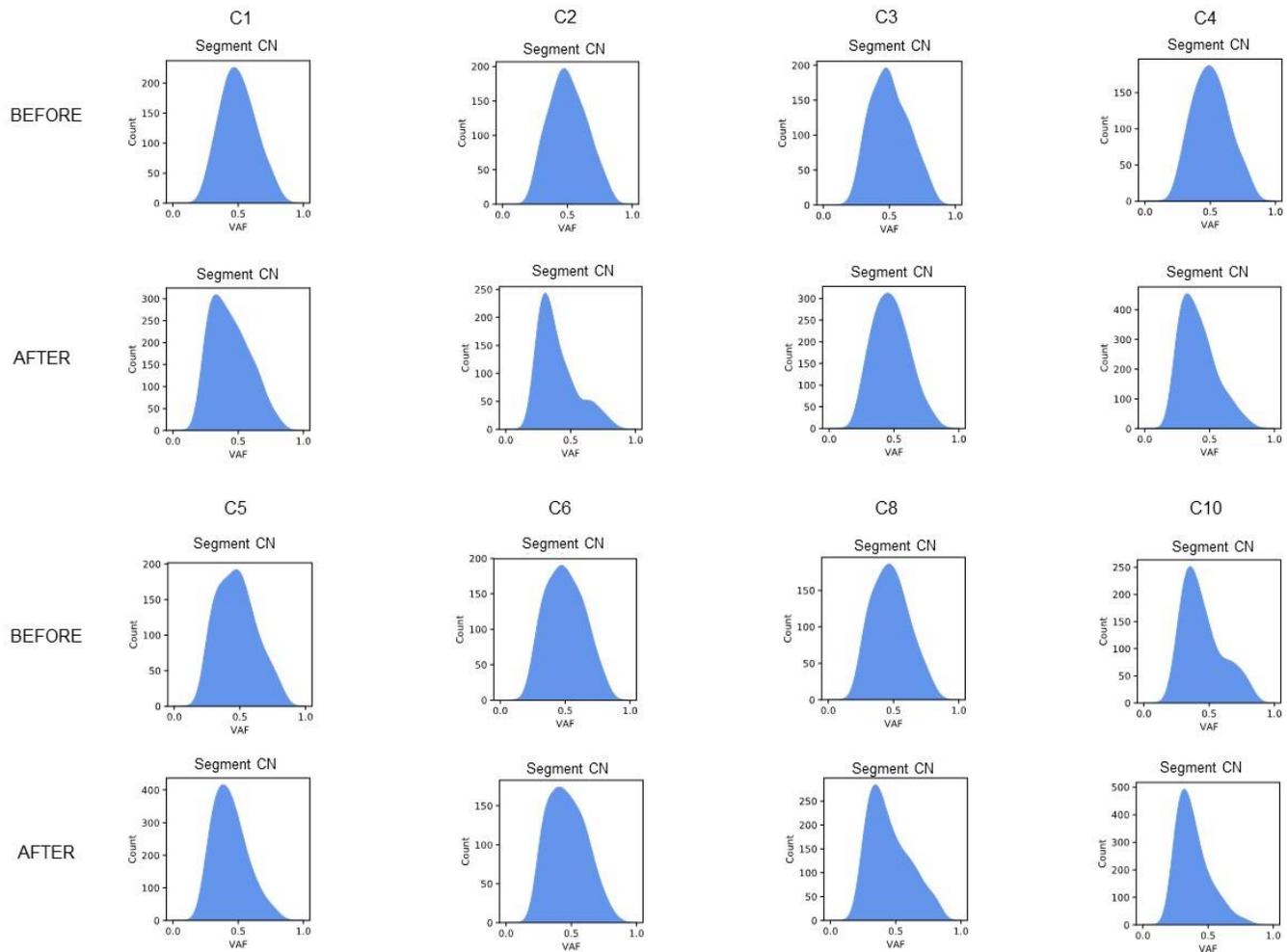
Overall, of the HAP1 WT<sup>DN-hTERT</sup> clones sampled before and after crisis, 75% were monoclonal at both time points. However, for the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones sampled before crisis, 50% were monoclonal, 50% were polyclonal. For HAP1 *POLQ*<sup>DN-hTERT</sup> clones sampled after crisis 75% of clones were polyclonal (**Figure 5.12B-C**). These data indicate that there is an increase in clonality as cells transit crisis in the absence of *POLQ*.

A

HAP1 WT<sup>DN-hTERT</sup> clones



**HAP1 POLQ<sup>DN-TERT</sup> clones**



B

HAP1 WT <sup>DN-hTERT</sup> BEFORE									
Clone	1	7	8	9	11	12	13	15	
Monoclonal	X	X		X	X		X	X	
Polyclonal			X			X			

HAP1 WT <sup>DN-hTERT</sup> AFTER									
Clone	1	7	8	9	11	12	13	15	
Monoclonal	X	X	X		X	X		X	
Polyclonal				X			X		

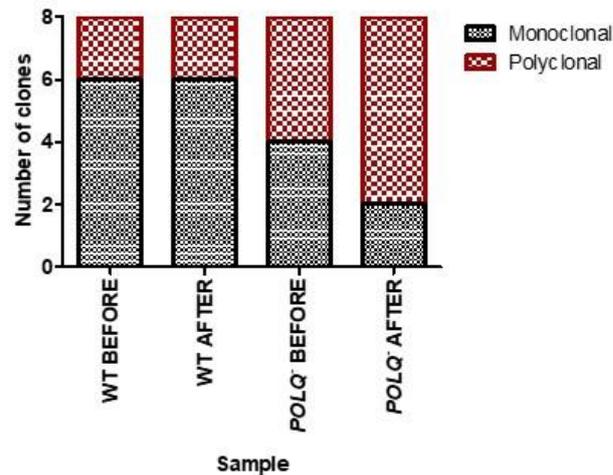
  

HAP1 POLQ <sup>:DN-hTERT</sup> BEFORE									
Clone	1	2	3	4	5	6	8	10	
Monoclonal		X	X	X		X			
Polyclonal	X				X		X	X	

HAP1 POLQ <sup>:DN-hTERT</sup> AFTER									
Clone	1	2	3	4	5	6	8	10	
Monoclonal			X			X			
Polyclonal	X	X		X	X		X	X	

C



**Figure 5.12: (A)** The clonality of the samples was assessed by analysing the VAF profile at diploid sites within the genome for HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> samples before and after crisis. Mono- or poly-clonality was manually inferred by comparing these profiles against theoretical distributions of mono or polyclonal cultures. **(B)** Tables showing clonality of HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> clones before and after crisis. **(C)** Stacked bar chart showing number of monoclonal or polyclonal HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> clones both before (passage 5-8) and after telomere-driven crisis (passage 32-46).

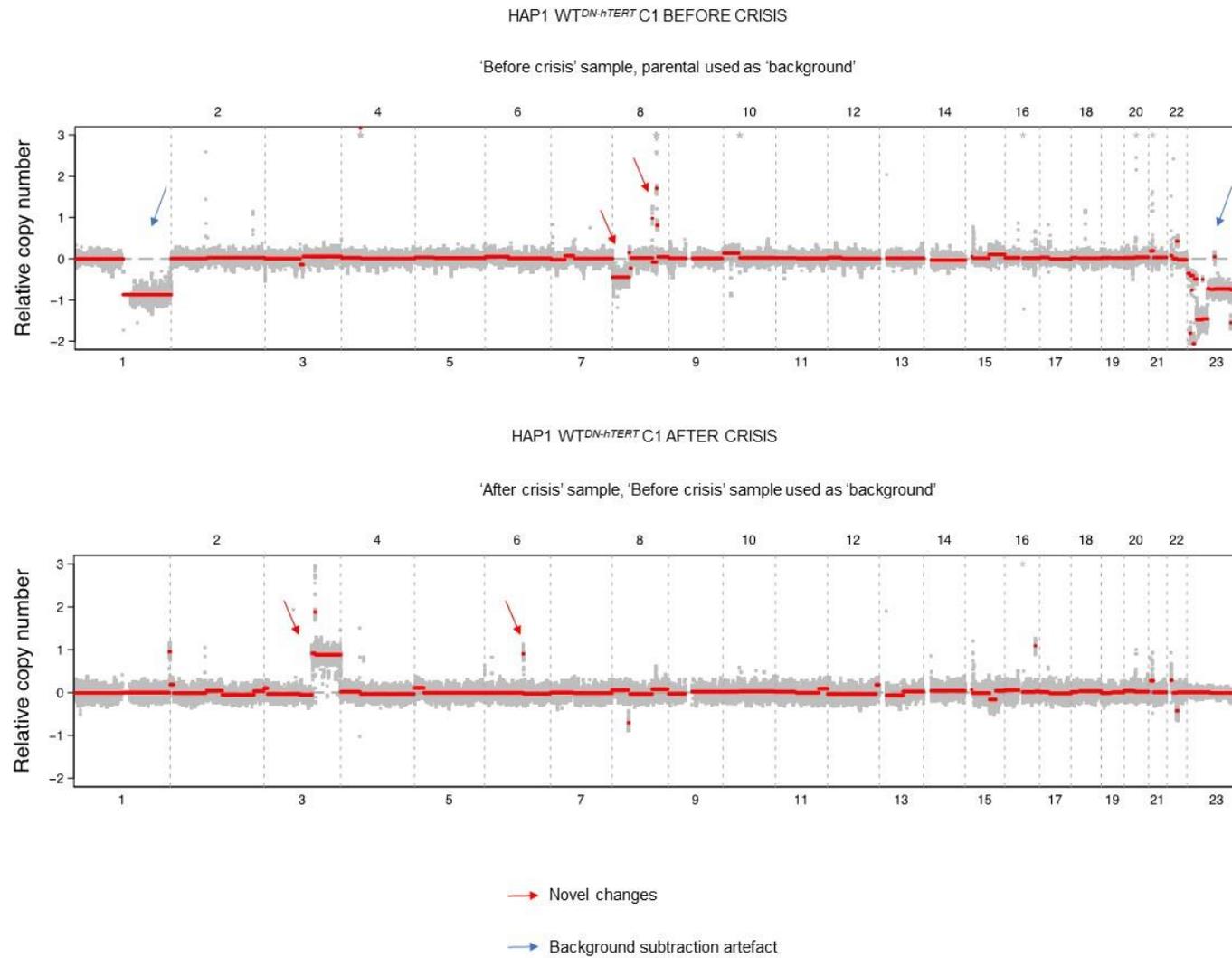
### Copy Number Analysis

Copy number variations (CNVs) are genomic alterations that result in an abnormal number of copies of chromosomal material and have been associated with a variety of human cancers (Shlien and Malkin 2009). Therefore, copy number was analysed in these cell lines to determine whether these genomic alterations arose during telomere-driven crisis. For copy number analysis, the read depth was normalised to the genome median and 2 'background' samples were used to identify changes before and after crisis. For changes that occurred before crisis, the parental background profile was utilised, whereas for changes that occurred after crisis the 'before crisis' background profile was utilised for each clone (**Figure 5.13A**). Several techniques were implemented to deal with artefacts (*Chapter 2*).

Large-scale CN variants (CNVs) from post-crisis clones were plotted as a heat map with gains in red and losses in blue, and segment means limited to the range (+1, -1) (**Figure 5.13B**). These data show that for the HAP1 WT<sup>DN-hTERT</sup> cell line, chromosome 3 is associated with 4.3-fold greater number of gains compared with the HAP1 POLQ<sup>-DN-hTERT</sup> cell line. Only one of these CNVs encompassed the POLQ locus.

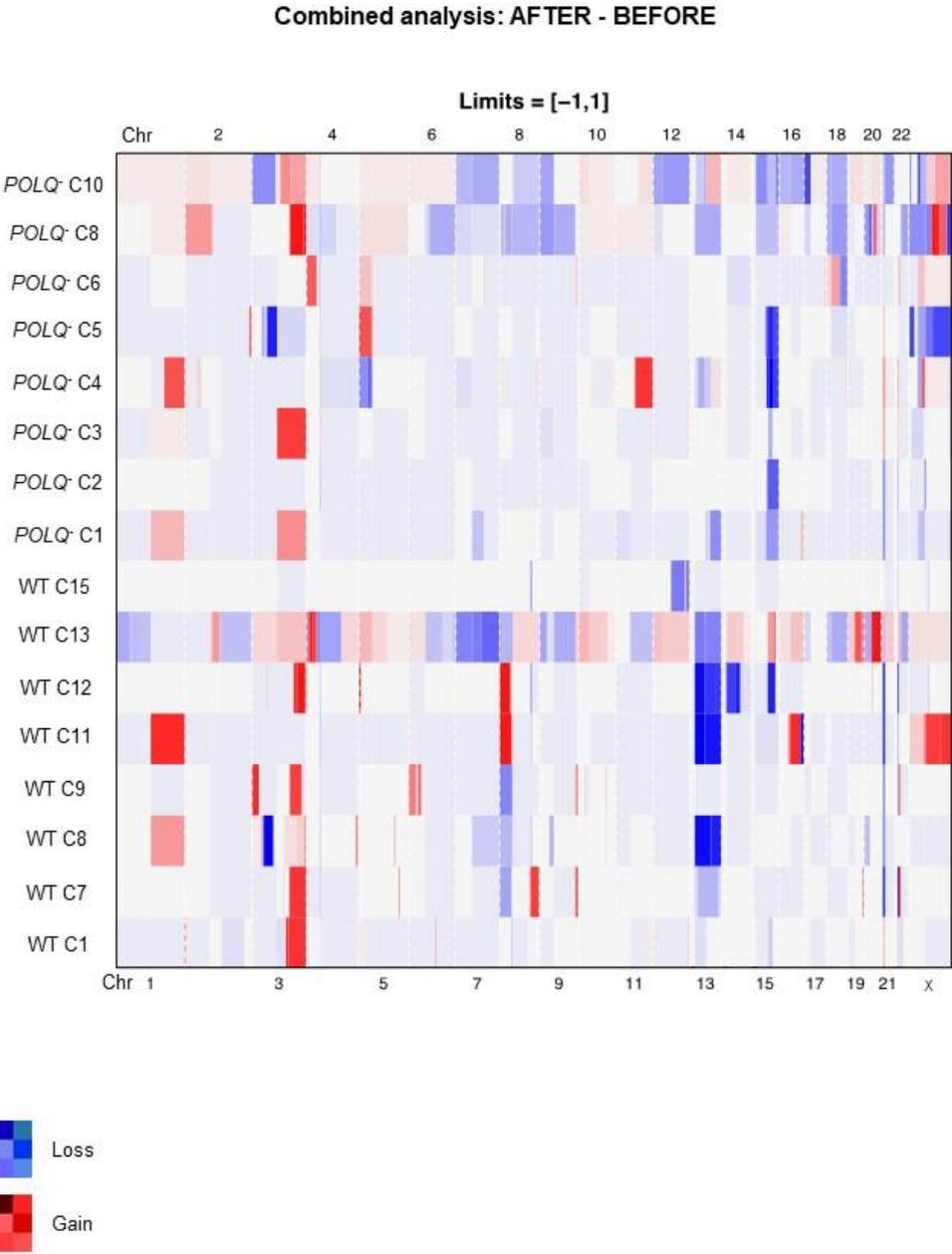
Whilst an increase in copy number gains and losses were observed for both cell lines after crisis, the only statistically significant increase recorded was for the CNV gains for the HAP1 WT<sup>DN-hTERT</sup> cell line ( $p < 0.05$ , Mann Whitney test). No significant differences in the number of copy number gains or losses were observed when comparing the two cell lines before and after crisis (**Figure 5.13C**).

A



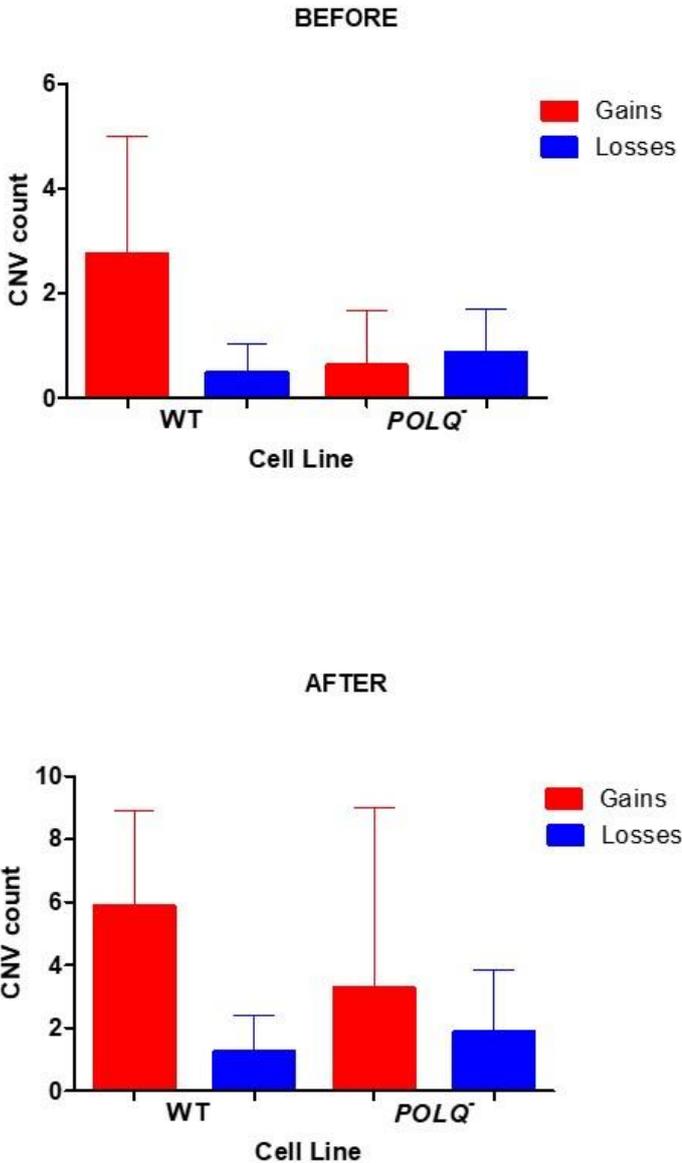
**Figure 5.13: (A)** Example of background sample subtraction for calling structural variants (TIDDIT software used).

B



**Figure 5.13: (B)** Genomic CN changes of HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ:<sup>DN-hTERT</sup> post-crisis clones are presented as normalized values relative to the before crisis clones within the limits of +1 gain and -1 loss.

C



**Figure 5.13: (C)** Bar charts showing the number of copy number gains and losses for HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>-/-DN-hTERT</sup> clones before (top) and after crisis (bottom). Error bars shown standard deviation and significance was determined using a Mann Whitney test.

### Structural Variants

Structural variants (SVs) are large ( $\geq 100$  bp) genomic alterations that can cause genomic instability that drives cancer progression (Hayes 2019). Therefore, SVs were analysed in these cell lines to ascertain whether these genomic alterations could have arisen during telomere-driven crisis. SVs were called using TIDDIT software (Eisfeldt et al. 2017). Default parameters were used and then filtered to keep unique SVs only. This was done using a pool of 'normal' samples. The parental samples were used as 'normal' samples to analyse the 'pre-crisis' samples, whereas the 'pre-crisis' samples were used as the 'normal' samples used to analyse 'post-crisis' samples. Each SV was cross checked against the pool to ensure no read-pairs supported the SV in any of the pool-samples.

The minimum variant size for all SVs was 100 bp and SVs were then categorised as deletions, duplications, inversions and translocations:

1. A deletion is defined as DNA sequence which is lost.
2. A duplication is defined as production of an additional copy of DNA.
3. An inversion is defined as a chromosome rearrangement in which a segment of a chromosome is reversed end to end.
4. A translocation is defined as a portion of a chromosome which has broken and reattached to a different chromosome.

For the HAP1  $WT^{DN-hTERT}$  clones, a total of 58 unique SVs were identified in pre-crisis samples. Following telomere-driven crisis, a 2.3-fold increase in unique SVs was observed in post-crisis samples, for which, a total of 133 SVs were identified. This represents a significant increase in the total number of unique SVs identified ( $p \leq 0.01$ , Mann Whitney test).

For the HAP1  $POLQ^{DN-hTERT}$  clones a total of 35 unique SVs were identified in pre-crisis samples compared to a total of 74 unique SVs identified for the post-crisis samples. This represents a significant 2.1-fold increase in SVs identified ( $p \leq 0.01$ ,

Mann Whitney test). Whilst there was no significant difference in the total number of SVs observed in pre-crisis samples for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>-DN-hTERT</sup> cell line, there was a significantly greater total number of unique SVs identified in post-crisis samples for the HAP1 WT<sup>DN-hTERT</sup> cell line compared to the HAP1 POLQ<sup>-DN-hTERT</sup> cell line ( $p \leq 0.01$ , Mann Whitney test) (**Figure 5.14A**).

### Pre-crisis samples

For the HAP1 WT<sup>DN-hTERT</sup> pre-crisis samples, the most common SVs identified were duplications (34%) and translocations (28%). Inversions and deletions accounted for 19% each of the total SVs identified. For the HAP1 POLQ<sup>-DN-hTERT</sup> pre-crisis samples, the most common SVs identified were also duplications (49%) and translocations (23%). Inversions accounted for 9% and deletions 20% of the total SVs identified. There were no significant differences in the proportions of different SVs identified between the HAP1 WT<sup>DN-hTERT</sup> pre-crisis samples and the HAP1 POLQ<sup>-DN-hTERT</sup> pre-crisis samples (**Figure 5.14B-C**).

### Post-crisis samples

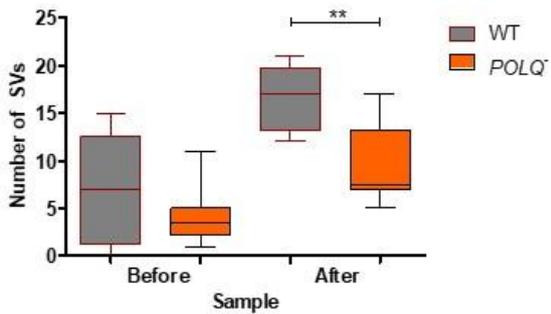
A significant increase in the number of duplications and translocations ( $p < 0.05$  and  $p < 0.01$  respectively, Mann Whitney test) was observed in post-crisis samples compared with pre-crisis samples for the HAP1 WT<sup>DN-hTERT</sup> cell line. Whilst an increase in the abundance of deletions and inversions was also observed, these were not statistically significant. Thus, as with the pre-crisis samples, the most common SVs identified for HAP1 WT<sup>DN-hTERT</sup> post-crisis samples were also duplications (39%) and translocations (35%). Inversions represented 16% whereas deletions accounted for 11% of the total number of SVs identified.

For the HAP1 POLQ<sup>-DN-hTERT</sup> cell line, a significant increase in the number of translocations was observed in post-crisis samples compared with pre-crisis samples

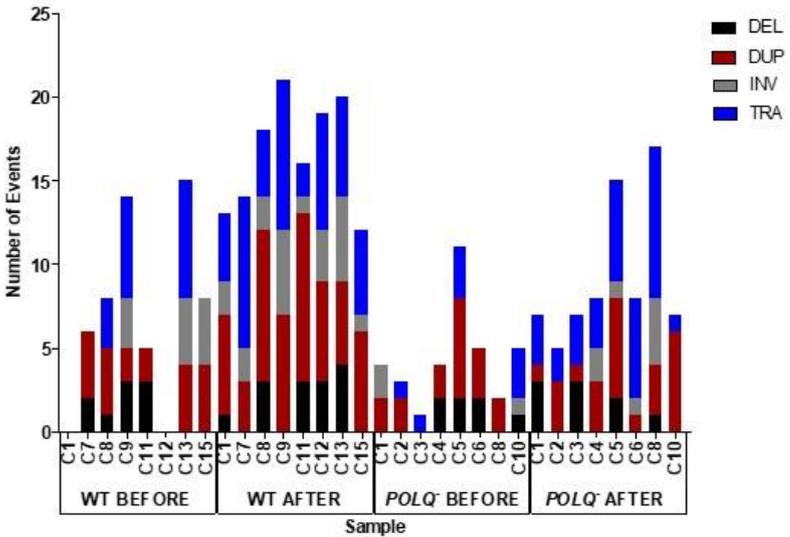
( $p < 0.05$ , Mann Whitney test). Although an increase in the number of duplications, deletions and inversions was also identified, these increases were not statistically significant. Therefore, for the post-crisis samples, the most common SVs identified were translocations (45%). Duplications represented 32% of the all SVs identified, thus, a decrease in the proportion of duplications was observed when compared with the pre-crisis samples. Inversions accounted for 11% and deletions accounted for 12% of all the SVs identified in the post-crisis samples. **(Figure 5.14D)**

Significantly more duplications and inversions were identified in HAP1  $WT^{DN-hTERT}$  post-crisis samples compared with HAP1  $POLQ^{-DN-hTERT}$  post-crisis samples ( $p < 0.05$ , Mann Whitney test). Despite the differences in the abundance of events, there was no significant difference in the proportion of the type of SVs when comparing the two cell lines **(Figure 5.14E)**.

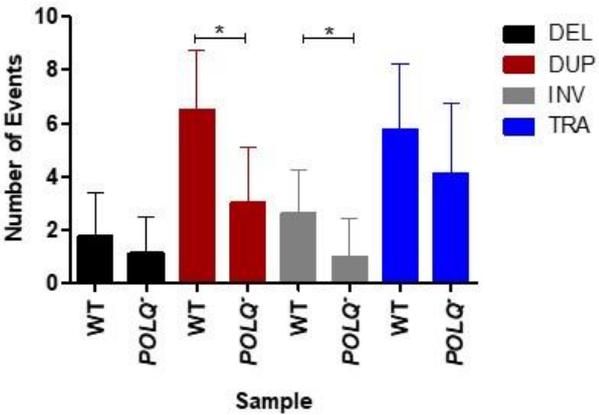
A



B



C





### Break site repair analysis

In order to analyse any break site repair features, SVs were assembled into contigs, (set of overlapping DNA segments) (Cleal et al. 2019) and any microhomology and insertions at break sites were quantified.

### Microhomology

For the HAP1 WT<sup>DN-hTERT</sup> pre-crisis clones, 64% of the SVs displayed microhomology as defined by 1 bp or more of common sequence. The mean length of the microhomology sequence for these clones was 4 bp. However, in post-crisis samples, a significant 2.3-fold increase in the abundance of SVs with microhomology was identified ( $p < 0.05$ , Mann Whitney test). However, no significant difference was seen when comparing the length of microhomology or the proportion of events with microhomology in pre-crisis and post-crisis HAP1 WT<sup>DN-hTERT</sup> clones.

For the HAP1 POLQ<sup>:DN-hTERT</sup> pre-crisis clones, 54% displayed microhomology. However, a significant 2.68-fold increase in the abundance of events displaying microhomology (69%) was observed in HAP1 POLQ<sup>:DN-hTERT</sup> post-crisis clones ( $p < 0.05$ , Mann Whitney test). The average length of the microhomology sequence detected for pre-crisis clones was 3.5 bp. No significant differences in the length of the microhomology detected or the proportion of events with microhomology were observed in pre-crisis and post-crisis HAP1 POLQ<sup>:DN-hTERT</sup> clones.

There was therefore no significant difference in the abundance ( $p = 0.2237$ , Mann Whitney test) or the proportion ( $p = 0.3894$ , Fisher's exact test) of SVs with microhomology or the length of the microhomology identified in pre-crisis clones when comparing the HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> cell lines.

Despite significant increases in the abundance of SVs that display microhomology at the break site in post-crisis samples compared to pre-crisis samples for both cell lines, the number of these events was significantly greater for the HAP1 WT<sup>DN-hTERT</sup> cell line ( $p < 0.05$ , Mann Whitney test) although no significant difference in the proportion of

SVs with microhomology was observed when comparing the cell lines ( $p=0.6488$ , Fisher's exact test). No significant difference was identified for the length of the microhomology detected in each cell line (**Figure 5.15A-B, E**).

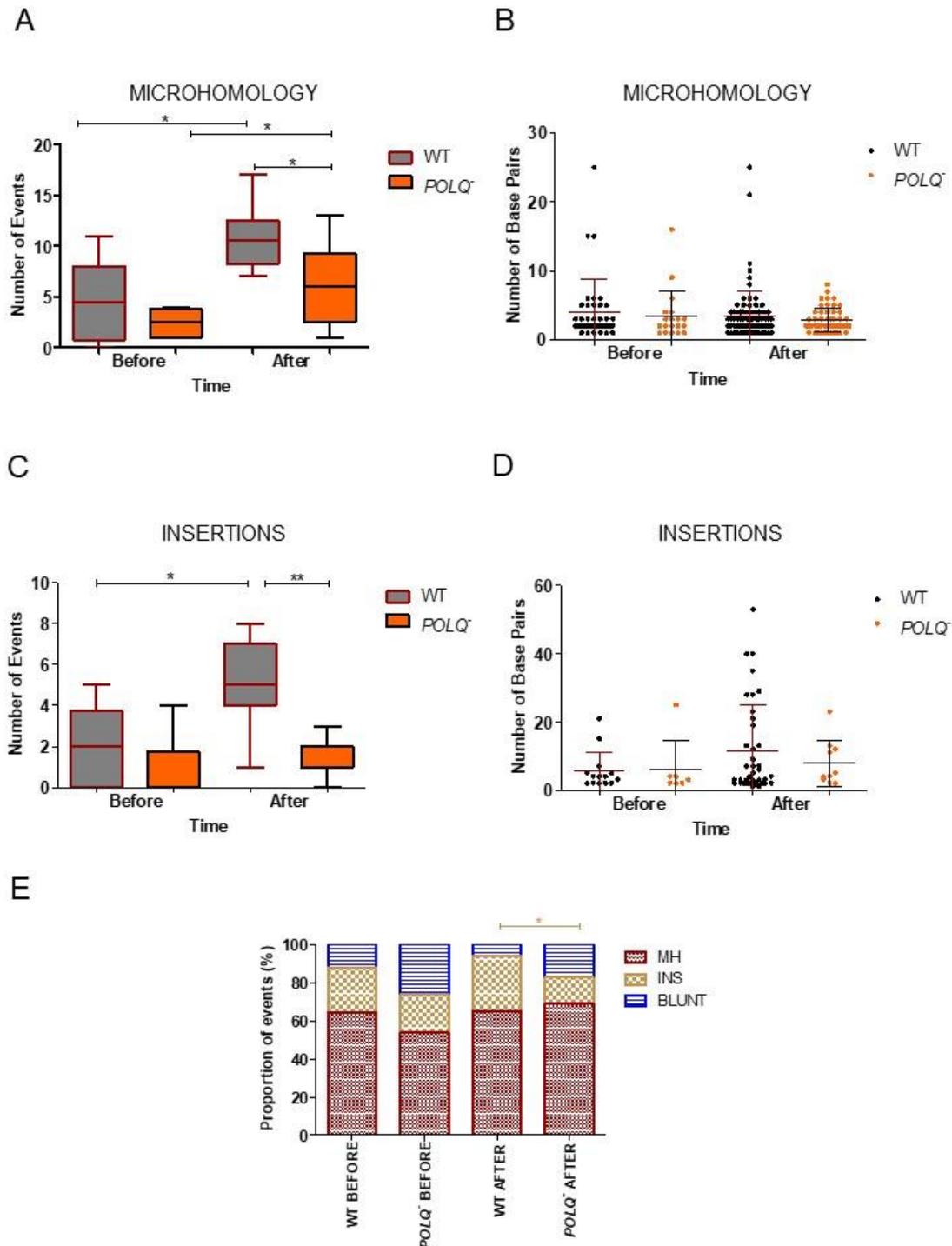
### Insertions

For the HAP1 WT<sup>DN-hTERT</sup> pre-crisis clones, 24% of the SVs identified had insertions at the break site. The average length of these insertions was 6 bp. A significant increase in the abundance of these events was observed in the post-crisis samples ( $p<0.05$ , Mann Whitney test), however, no significant differences were observed for the proportions of these events (29%) or the length of the insertions detected. However, a greater variation in the range of insertion lengths recorded for the post-crisis clones compared with the pre-crisis clones was observed. This is reflected by the difference in the standard deviation values of 5.39 bp for pre-crisis samples compared with 13.33 bp for post-crisis samples.

For the HAP1 POLQ<sup>-DN-hTERT</sup> pre-crisis clones, 20% of the SVs identified had insertions at the break site and the average length of these insertions was 6 bp. Despite an increase in the abundance of SVs with insertions and an increase in the length of the insertions in the post-crisis samples, these increases were not statistically significant. Likewise, no change in the range of insertion lengths was detected for pre-crisis (7.80 bp) compared with post-crisis samples (6.43 bp). A decrease in the proportion of these events (14%) was observed for the post-crisis samples compared to the pre-crisis samples although this was also not statistically significant.

Despite an increase in the abundance of SVs that had insertions at the break site in post-crisis samples for both cell lines, the number of these events was significantly greater for the HAP1 WT<sup>DN-hTERT</sup> cell line ( $p\leq 0.01$ , Mann Whitney test). SVs with insertions represented a significantly greater proportion of events for the HAP1 WT<sup>DN-</sup>

*hTERT* post-crisis clones compared with the HAP1 *POLQ<sup>-DN-hTERT</sup>* post-crisis clones ( $p < 0.05$ , Fisher's exact test) (**Figure 5.15C-E**).



**Figure 5.15:** (A) Box and Whisker plot showing the number of SVs with microhomology for HAP1<sup>DN-hTERT</sup> clones before and after crisis. Error bars indicate minimum and maximum values and significance was determined using a Mann Whitney test. (B) Scatter graph showing length of microhomology detected for SVs in HAP1<sup>DN-hTERT</sup> clones before and after crisis. (C) Box and Whisker plot showing the number of SVs with insertions for HAP1<sup>DN-hTERT</sup> clones before and after crisis. Error bars indicate minimum and maximum values and significance was determined using a Mann Whitney test. (D) Scatter graph showing length of insertions detected for SVs in HAP1<sup>DN-hTERT</sup> clones before and after crisis. Error bars indicate standard deviation and significance was determined using a Mann Whitney test. (E) Stacked bar chart showing the proportion of SVs with microhomology, insertions or blunt junctions for HAP1<sup>DN-hTERT</sup> cell lines before and after crisis. Significance was determined using a Fisher's exact test.

### Comparing genomic fusion events with SVs

Genomic fusion events recorded for 5 HAP1 WT<sup>DN-hTERT</sup> and 5 HAP1 POLQ<sup>:DN-hTERT</sup> clones were intersected with SV break sites recorded for the same clones from the whole genome sequencing data to determine whether any of the events coincided. SV break sites were intersected with fusion intervals by expanding fusion intervals by 1, 10 and 25 Mb (e.g. 1 Mb was subtracted from start coordinate and 1 Mb was added to end coordinate). These expanded intervals were intersected with the SV break sites using BEDtools (Quinlan and Hall 2010), and the statistical significance of overlaps was assessed using BEDtools' Fisher function which implements a Fisher's exact test to identify whether there were significantly more or less intersecting overlaps than estimated. The data analysis using BEDtools was performed by Dr Kez Cleal, Cardiff University.

As a 4-fold greater abundance of genomic fusion events was recorded for the HAP1 POLQ<sup>:DN-hTERT</sup> compared with the HAP1 WT<sup>DN-hTERT</sup> cell line, a 1.7-fold greater number of fusion events (25 Mb) that overlapped with SV break sites were also recorded for the HAP1 POLQ<sup>:DN-hTERT</sup> compared with the HAP1 WT<sup>DN-hTERT</sup> cell line.

For both cell lines, a greater number of fusion events that overlapped with SV break sites was identified as the fusion intervals increased from 1 Mb to 25 Mb. A 6.5-fold and 10-fold increase in the abundance of events was observed when comparing events detected at 1 Mb and 10 Mb fusion intervals for the HAP1 WT<sup>DN-hTERT</sup> and HAP1 POLQ<sup>:DN-hTERT</sup> cell lines, respectively. Likewise, a 1.8-fold and 2-fold increase in the number of events was recorded for 25 Mb compared with 10 Mb fusion intervals.

At 1 Mb fusion intervals only 2 fusion events that overlapped with SV break sites were identified for both cell lines. However, at 10 Mb fusion intervals, 13 fusion events were recorded which overlapped with 12 SVs for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with 20 events for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line (**Figure 5.16**). For the HAP1 WT<sup>DN-hTERT</sup> cell line, the most SV-coinciding fusion events were associated with

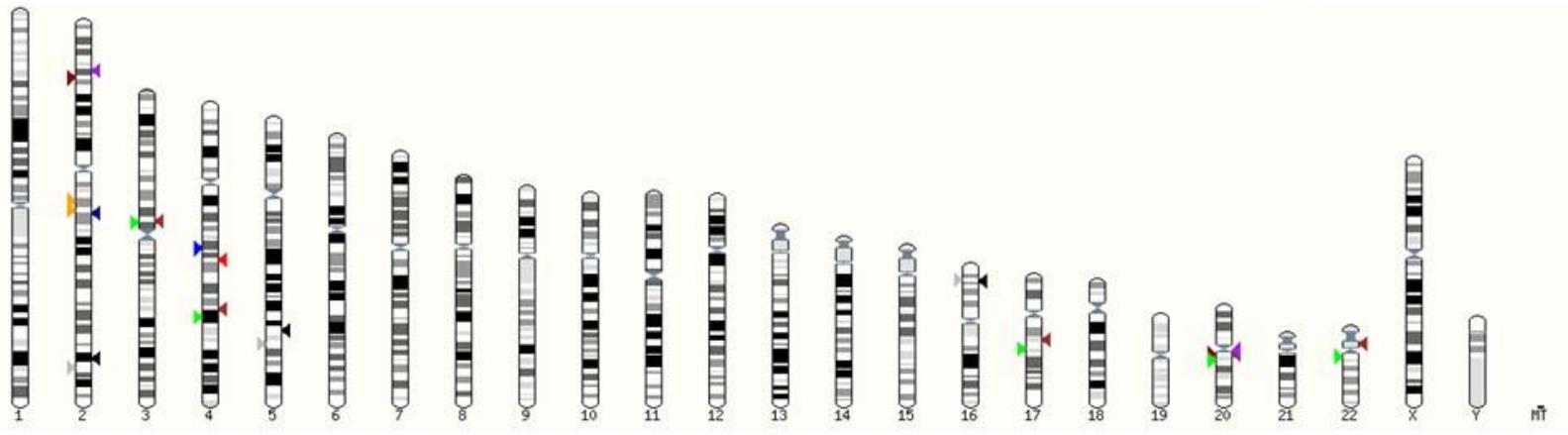
chromosome 2 (4 fusion events that overlapped with 3 SVs) whereas the most SV-coinciding fusion events identified for the HAP1 *POLQ<sup>-DN-hTERT</sup>* cell line were associated with chromosome 1 (5 fusion events that overlapped with 3 SVs). For both cell lines, chromosome 1 was involved in the most genomic fusion events overall compared with other chromosomes (10 and 35 genomic fusion events).

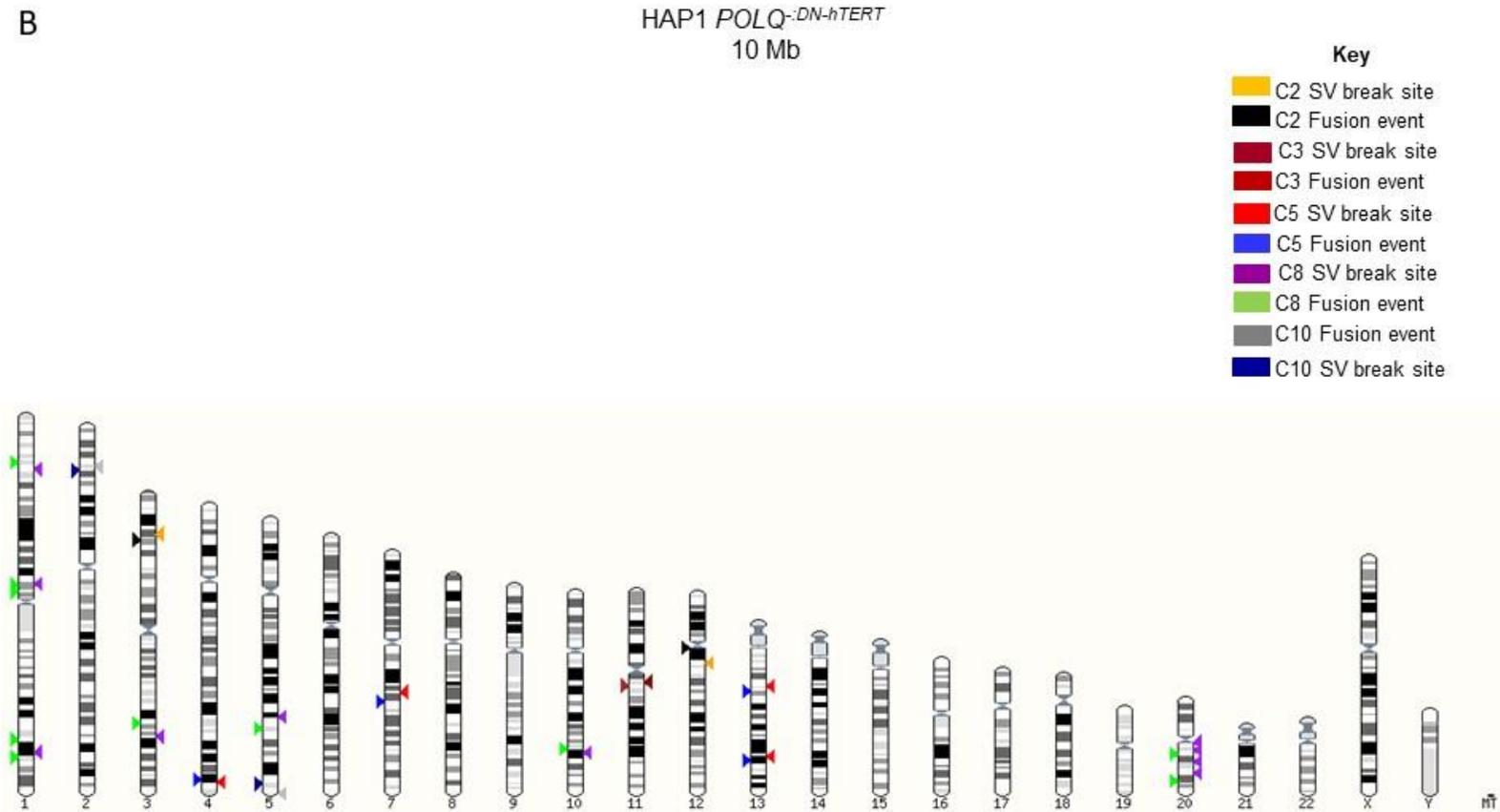
Similarly, for both cell lines, SV coincidence was associated with elevated fusion frequency even after adjusting for chromosome size. For example, a 6.8-fold and 3.2-fold increase in genomic fusion events was observed for chromosome 22 and chromosome 20 compared with predicted fusion frequency based on individual chromosome size for the HAP1 *WT<sup>DN-hTERT</sup>* cell line. Fusion events that occurred for these chromosomes also overlap with SV break sites. Likewise, for the HAP1 *POLQ<sup>-DN-hTERT</sup>* cell line, a 3.3-fold increase in genomic fusion events was recorded for chromosome 20 compared with predicted fusion frequency based on its size and 2 of these fusion events overlapped with 4 SV break sites.

Therefore, for both cell lines, the abundance of fusion events that overlapped with SVs is associated with the frequency of genomic fusion events recorded. Interestingly for the HAP1 *WT<sup>DN-hTERT</sup>* cell line, no fusion events that overlapped with SV break sites were observed for chromosome 1 despite its size and for the HAP1 *POLQ<sup>-DN-hTERT</sup>* cell line no fusion events overlapped with SVs on chromosomes 19 and 22 even though these chromosomes had 6.37-fold and 6.05-fold greater numbers of fusion events than predicted based on chromosome size. This may be because this analysis is focused on SVs that arise during crisis, therefore, any underlying instabilities that may give rise to fusions have been normalised.

A

HAP1 WT<sup>DN-hTERT</sup>  
10 Mb





**Figure 5.16:** Karyotype diagram (Ensembl) showing genomic fusion events at 10 Mb intervals (left arrows) overlapping with SV break sites (right arrows) observed for HAP1 WT<sup>DN-hTERT</sup> (**A**) and HAP1 *POLQ*<sup>-DN-hTERT</sup> clones (**B**).

A total of 23 fusion events that overlapped with 23 SV break sites were recorded for the HAP1 WT<sup>DN-hTERT</sup> cell line at 25 Mb fusion intervals compared to 40 fusion events that overlapped with 37 SVs for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line (**Figure 5.17**).

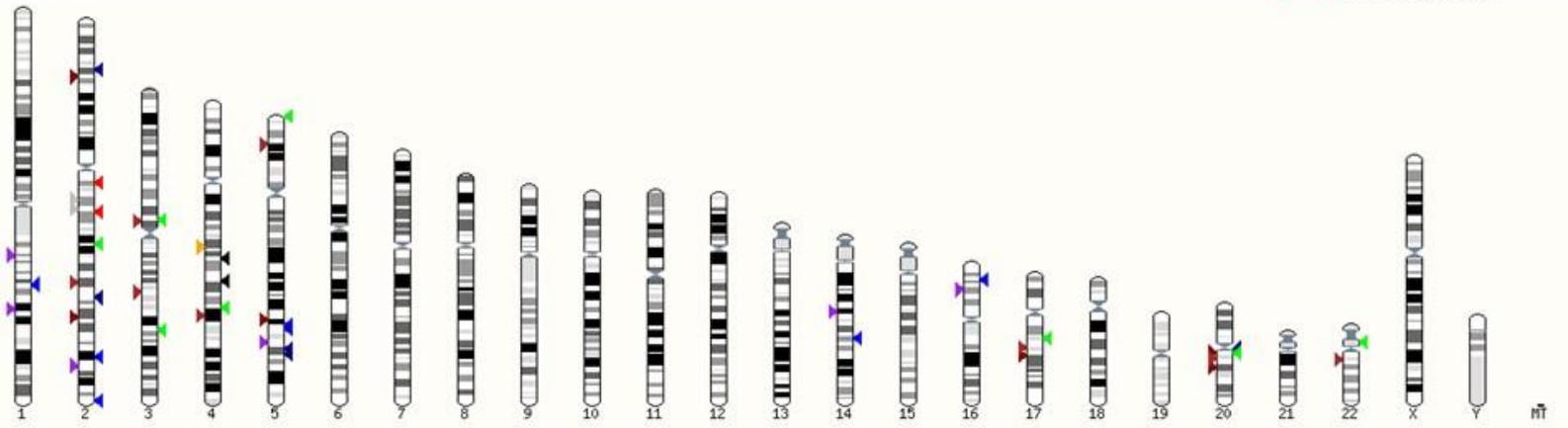
As with events recorded at 10 Mb fusion intervals, chromosome 2 was also involved in more events (6 fusion events that overlapped with 7 SV break sites) compared with other chromosomes at 25 Mb fusion intervals for the HAP1 WT<sup>DN-hTERT</sup> cell line. Likewise, chromosome 1 was also involved in more events compared with other chromosomes for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line (9 fusion events that overlapped with 4 SVs). Interestingly, 8 of the 9 fusion events located on chromosome 1 were associated with C8. Furthermore, 19 fusion events that overlapped with 16 SV break sites were identified for this clone. This accounts for 48% of all fusion events that overlapped with SV break sites for this cell line. Whilst the greatest number of genomic fusion events was observed for this clone, the same number of fusion events (96) were also recorded for HAP1 POLQ<sup>:DN-hTERT</sup> C2, however, only 6 fusion events that overlapped with 5 SV break sites were observed for this clone. C10 had the lowest genomic fusion frequency (43 events) but had a similar number of fusion events that overlapped with SVs break sites (5) as C2. Thus, an increased frequency of fusion events that overlapped with SV break sites was observed for C8 compared with clones with similar genomic fusion frequencies.

However, no statistically significant overlaps were identified for any of the samples. There were, however, some samples that had significantly less overlaps than expected (Fisher's exact tests): HAP1 WT<sup>DN-hTERT</sup> C11 (10 Mb fusion intervals,  $p < 0.05$ ) and C12 (25 Mb fusion intervals,  $p \leq 0.01$ ) as well as HAP1 POLQ<sup>:DN-hTERT</sup> C3 (25 Mb fusion intervals,  $p \leq 0.001$ ) and C5 (25 Mb fusion intervals,  $p \leq 0.01$ ).

A

HAP1 WT<sup>DN-hTERT</sup>  
25 Mb

- Key**
- C1 SV break site
  - C1 Fusion event
  - C7 SV break site
  - C7 Fusion event
  - C8 Fusion event
  - C8 SV break site
  - C11 SV break site
  - C11 Fusion event
  - C12 SV break site
  - C12 Fusion event

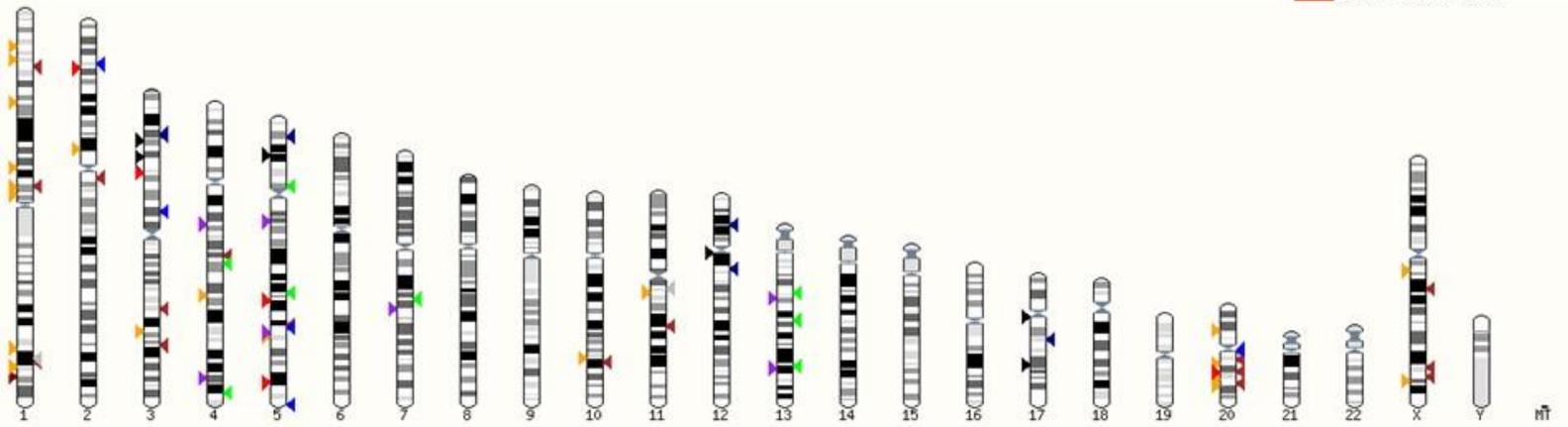


B

HAP1 *POLQ*<sup>-DN-hTERT</sup>  
25 Mb

Key

- C2 SV break site
- C2 Fusion event
- C3 SV break site
- C3 Fusion event
- C5 SV break site
- C5 Fusion event
- C8 SV break site
- C8 Fusion event
- C10 SV break site
- C10 Fusion event



**Figure 5.17:** Karyotype diagram (Ensembl) showing genomic fusion events at 25 Mb intervals (left arrows) overlapping with SV break sites (right arrows) observed for HAP1 WT<sup>DN-hTERT</sup> (A) and HAP1 *POLQ*<sup>-DN-hTERT</sup> clones (B).

## 5.5 DISCUSSION

### 5.5.1 Fusion Amplicon Sequencing Data

Following Illumina HiSeq4000 paired-end next-generation sequencing of telomere fusion amplicons from HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>-DN-hTERT</sup> clones transiting a telomere driven crisis, a 2.6-fold greater number of fusion events (raw data) was observed for the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line compared with the HAP1 WT<sup>DN-hTERT</sup> cell line. These telomere fusion events were categorised as genomic-telomere (fusion between subtelomere and other genomic location), 16p-21q inter-chromosomal (fusion between 16p and 21q family subtelomeric regions) or '16p' / '21q' fusion events (fusion between 16p/16p or 21q/21q family subtelomeric regions).

#### **Disruption of *POLQ* associated with increased proportion of inter-chromosomal fusion events**

Regardless of overall fusion abundance (as samples were not normalised), the HAP1 WT<sup>DN-hTERT</sup> cell line had a significantly greater proportion of '16p' / '21q' events (80%) and a significantly smaller proportion of both genomic (18%) and inter-chromosomal (2%) fusion events compared with the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line (38%, 29%, 33%).

These data show that the disruption of *POLQ* results in an increase in the proportion of inter-chromosomal fusion events. It has previously been demonstrated that A-NHEJ is associated with intra-chromosomal fusion events whilst C-NHEJ is associated with inter-chromosomal fusion events (Jones et al. 2014). The increased proportion of genomic and 16p-21q inter-chromosomal fusion events observed for the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line could therefore be due to the disruption of *POLQ* and the subsequent upregulation of other DSB repair mechanisms such as C-NHEJ. The reduction in the proportion of genomic and 16p-21q inter-chromosomal fusion events for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line

may be due to the lower incidence of DSBs as both A-NHEJ, HR and C-NHEJ are intact. It has been demonstrated that when the overall incidence of DSBs is low, HR contributes to the repair of ~50% of DSBs. However, following an increase in DSB frequency and genome instability (via irradiation of cells), repair pathways dependent on resection, including HR and A-NHEJ, were suppressed. Therefore, an upregulation in SSA-mediated repair and C-NHEJ (not dependent on resection) was observed following an elevated incidence of DSBs (Mladenov et al. 2019). Thus, the increase in 16p-21q inter-chromosomal and genomic telomere fusion events in the HAP1 *POLQ<sup>:DN-hTERT</sup>* cell line may be due to an elevated incidence of unresolved DSBs as A-NHEJ is inhibited which raises the overall DSB frequency and genomic instability to a threshold at which HR is suppressed.

The threshold for HR/A-NHEJ suppression may have already been surpassed in these cell lines as the incidence of DSB increased as a function of telomere shortening during telomere-driven crisis. It is therefore plausible that the HAP1 *POLQ<sup>:DN-hTERT</sup>* cell line is disproportionately dependent on C-NHEJ, causing an increased abundance of inter-chromosomal and genomic fusion events compared with the HAP1 *WT<sup>DN-hTERT</sup>* cell line. To determine whether HR is suppressed, RAD51 expression could be analysed for these clones undergoing telomere-driven crisis.

Although a significantly greater frequency of SVs was detected for the HAP1 *WT<sup>DN-hTERT</sup>* cell line compared with the HAP1 *POLQ<sup>:DN-hTERT</sup>* cell line, the potential decrease in sensitivity for detecting SVs for the HAP1 *POLQ<sup>:DN-hTERT</sup>* cell line should be considered. This drop in sensitivity may be due to the discrepancy in clonality between the two cell lines as a greater proportion of HAP1 *POLQ<sup>:DN-hTERT</sup>* clones were polyclonal following crisis (75%) compared with HAP1 *WT<sup>DN-hTERT</sup>* clones (25%). An increased dependence on C-NHEJ and subsequent increase in inter-chromosomal fusion events for the HAP1 *POLQ<sup>:DN-hTERT</sup>* clones may have contributed to the increased polyclonality of the cell line due to increased genomic heterogeneity. This

discrepancy in the proportion of polyclonal clones between the cell lines could provide an explanation as to how the POLQ-deficient cell lines escape crisis more readily than the WT cell lines. This is consistent with the growth curves obtained for clonal populations transiting telomere-driven crisis (*Chapter 3*). Together, these data indicate that the transit through crisis was more rapid in the absence of POLQ. This may be due to these POLQ-deficient cells having a greater mutational load and genetic heterogeneity on which clonal selection can operate.

To test this hypothesis, whole genome sequencing of HCT116 WT<sup>DN-hTERT</sup> and HCT116 POLQ<sup>-/-:DN-hTERT</sup> clones before and after crisis could be performed to determine the abundance of inter-chromosomal fusion events and the clonality of these cell lines as the HCT116 POLQ<sup>-/-:DN-hTERT</sup> clones also escaped crisis more readily than the HCT116 WT<sup>DN-hTERT</sup> clones (*Chapter 3*).

Since A-NHEJ is characterised by insertions and microhomology at DSB sites, fusion junctions were analysed to determine whether there were any differences in the abundance of microhomology and insertions at fusion junctions when comparing the two cell lines. Differences in the abundance of these features could suggest that different repair mechanisms had been utilised to facilitate telomere fusion events for these two HAP1 cell lines.

### **Fusion junction features: Microhomology**

No significant differences in the proportions of fusion junctions with microhomology were observed when comparing genomic fusion events for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>-/-:DN-hTERT</sup> cell line. However, the length of the microhomology and the range of distributions detected was significantly longer in the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>-/-:DN-hTERT</sup> cell line.

A greater proportion of fusion junctions with microhomology were observed for the HAP1 POLQ<sup>-/-:DN-hTERT</sup> cell line for both 16p-21q inter-chromosomal and '16p' / '21q'

fusion events compared to the HAP1 WT<sup>DN-hTERT</sup> cell line. However, there was no significant difference in the length of the microhomology observed for these events when comparing the two cell lines.

### **Fusion junction features: Insertions**

No significant differences in the proportions of fusion junctions with insertions were observed when comparing genomic and 16p-21q inter-chromosomal fusion events for the HAP1 WT<sup>DN-hTERT</sup> cell line with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line. However, a significantly greater proportion of fusion junctions with insertions was observed when comparing '16p' / '21q' events for the HAP1 WT<sup>DN-hTERT</sup> cell line with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line. For all event types (genomic, 16p-21q, '16p' / '21q'), the lengths of the insertions and the range in distributions observed were greater for the HAP1 WT<sup>DN-hTERT</sup> cell line.

These data may suggest that the same repair mechanisms are being utilised for both cell lines, with the exception of a few events in the HAP1 WT<sup>DN-hTERT</sup> cell line which give rise to longer microhomology and insertions at break sites. These events could be a result of single-strand break repair (SSBR) or base excision repair (BER) although increased microhomology and insertions are a feature of repair mediated by POLQ (Kent et al. 2015). Thus, in the HAP1 WT<sup>DN-hTERT</sup> cell line, DSB repair is primarily performed via mechanisms such as C-NHEJ (as with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line) but this cell line also has the capacity for A-NHEJ via activity of POLQ. This is consistent with previously published data that suggests that C-NHEJ is the dominant form of NHEJ if both pathways are intact and A-NHEJ activity increases if C-NHEJ is inhibited (Wang et al. 2003).

### **Genomic fusion events: chromosome and gene involvement**

Increased fusion frequency was associated with smaller chromosomes when actual fusion frequencies were compared with predicted fusion frequencies based on

chromosome size for both cell lines. Chromosomes 20 and 22 had the greatest abundance of fusion events relative to chromosome size for the HAP1 WT<sup>DN-hTERT</sup> cell line. Similarly, chromosomes 19, 22, 20 and 17 had the greatest abundance of fusion events relative to chromosome size for the HAP1 *POLQ*<sup>DN-hTERT</sup> cell line.

HAP1 cells (derived from the KBM7 CML line) have the Philadelphia translocation (between chromosome 9 and 22 and between chromosome 15 and 19 in initial cell line). Therefore, the prevalence of fusion events involving chromosomes 19 and 22 may be due to the genetic profile of the HAP1 cell lines.

In addition to the chromosomes involved, genomic fusion junctions were mapped to determine whether any genomic telomere fusion events overlapped with genes to evaluate the potential impact of these telomere fusion events on the genome more widely. Telomere fusions involving genes may disrupt gene function and thus cause substantial phenotypic changes. For both cell lines the frequency of genomic fusions that overlapped with genes was higher than expected based on the average gene content (43.4%) of the GRCh38.p13 reference human genome. This reflects previously published data which demonstrated that inter-chromosomal fusion junctions were significantly more likely to occur within genes. Likewise, this also reflects data which suggest that chromosomal breakpoints observed in cancer are preferentially located in genic regions (Liddiard et al. 2016).

The genes involved in telomere fusion events were associated with the same molecular functions, cellular and metabolic processes and protein classes for both cell lines. Notably, genes associated with the Wnt signalling pathway were identified for both cell lines (WT=2.3% and *POLQ* =5% of all genes identified in genomic fusions). This is similar to previously published data which showed an enrichment for telomere fusions that overlapped with genes containing binding motifs for Wnt pathway. This was observed for HCT116 cell lines under different stress conditions (Liddiard et al. 2016).

Both cell lines showed increased fusion events at smaller chromosomes and fusion events overlapped with genes involved in the same molecular functions, cellular and metabolic processes and protein classes for both cell lines. The observation that smaller chromosomes are involved in disproportionately more fusion events may be due to gene density. This conforms with data which suggests that some smaller chromosomes have greater gene densities. For example, chromosome 19 and 22 have relatively large gene densities (Sehgal et al. 2014).

Therefore, the formation of fusion events may be influenced by replication timing as gene rich regions are often early replicating regions (Donaldson and Nieduszynski 2019). The fusogenic potential of genomic loci may therefore be associated with euchromatin which is more accessible than heterochromatin. These data also conform with previously published data which suggests that most DSBs within gene-rich euchromatin are repaired by C-NHEJ (Muraki et al. 2015). Likewise, additional data suggests that chromosomes with high gene density are preferentially repaired in human cells (Surrallés et al. 1997).

If DNA replication is associated with repair, these gene-rich loci may present opportunity for rapid DSB repair via replication-independent NHEJ fusion. This may be due to increased substrate availability at genomic loci where repair and replication enzymes are already colocalised. Furthermore, cell cycle regulation of template processing and availability may also have an impact on the abundance and mechanism of telomere fusions (Escribano-Díaz et al. 2013).

### 5.5.2 Whole Genome Sequencing of HAP1<sup>DN-hTERT</sup> clones

HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>DN-hTERT</sup> clones were sampled before and after crisis for whole genome paired-end sequencing to determine the potential contribution of telomere fusions to the global genomic instability occurring in crisis.

### Structural Variants

A significant increase in the abundance of SVs was detected for samples taken after crisis compared with samples taken before crisis for both cell lines. However, a significantly greater frequency of SVs was detected for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line. This is unexpected since a significantly greater frequency of fusion events was observed for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line compared with the HAP1 WT<sup>DN-hTERT</sup> cell line. These data should be analysed whilst considering the discrepancy in clonality between the two cell lines and therefore the potential drop in sensitivity for detecting SVs for the HAP1 POLQ<sup>:DN-hTERT</sup> cell line.

No significant differences were observed when comparing the proportions of different SVs (deletions, duplications, inversions and translocations) for the two cell lines. Although a significantly greater proportion of SVs with insertions at the break site were observed for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line, no significant differences in the length of insertions were observed whilst comparing the two cell lines. Similarly, no significant difference in the proportion of SVs with microhomology or the length of the microhomology at the break site was observed when comparing the cell lines.

These data reflect what was observed for the fusion sequencing data and thus indicate a role for C-NHEJ in telomere fusions for both cell lines as no significant differences in the use of microhomology were observed for the two cell lines, suggesting a common repair mechanism was utilised. The difference in the proportion of SVs with insertions for the HAP1 WT<sup>DN-hTERT</sup> cell line compared with the HAP1 POLQ<sup>:DN-hTERT</sup> cell line could suggest a role for A-NHEJ (which is characterised by insertions at break sites) in a few events for the HAP1 WT<sup>DN-hTERT</sup> cell line only.

### Fusion events and SVs

Genomic fusion events were intersected with SV break sites to determine whether any of the events were coincident. A 1.7-fold increase in fusion events that overlapped with SV break sites were observed at 25 Mb fusion intervals for the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line compared with the HAP1 *WT*<sup>DN-hTERT</sup> cell line. For both cell lines the abundance of fusion events that overlapped with SVs is associated with the frequency of genomic fusion events recorded. These data suggest that genomic fusion events could give rise to the structural variants observed. Therefore, the telomere fusion events observed may drive genome instability and clonal evolution of cancer cells. This is supported by data which demonstrates increased genomic instability in immortalised cancer cell lines which have also reactivated telomerase activity following telomere-driven crisis (Maciejowski and De Lange 2017).

### 5.5.3 Conclusions

Taken together, these data indicate that the disruption of *POLQ* is associated with an increase in genomic and 16p-21q inter-chromosomal fusion events during telomere-driven crisis. The telomere fusion events observed were likely due to the use of C-NHEJ (which was assumed to be fully functional in both cell lines). However, some fusion events observed for the HAP1 *WT*<sup>DN-hTERT</sup> cell line may have resulted from the activity of the A-NHEJ pathway based on increased microhomology and insertion lengths which characterise this repair mechanism. It is possible that the increased frequency of genomic and inter-chromosomal fusion events for the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line was due to the upregulation of C-NHEJ via the suppression of HR and A-NHEJ as a result of increased DSB frequency. To investigate HR activity, RAD51 foci could be analysed as a functional biomarker of HR in these cells. It is feasible that this C-NHEJ may have also contributed to the incidence of SVs and changes in clonality as a result of telomere fusion events. It was not possible to determine the frequency of intra-chromosomal fusion events from these data, so the frequency of

intra-chromosomal fusion events associated with the downregulation of A-NHEJ in the HAP1 *POLQ<sup>-DN-hTERT</sup>* cell line could not be confirmed. However, previous data has demonstrated both the inhibition of A-NHEJ in the absence of POLQ (Mateos-Gomez et al. 2015) and the association of A-NHEJ with intra-chromosomal telomere fusion events (Jones et al. 2014).

These data suggest that the disruption of *POLQ* is associated with increased capacity for escape from crisis due to increased genomic heterogeneity. It has previously been suggested that *POLQ* represents a valid therapeutic target for inhibition in human cancers that are HR-deficient. However, these data could suggest that instead of achieving a synthetic lethal approach by combining *POLQ* inhibition and chemotherapy, these conditions could favour increased survival of cancer cells via upregulation of C-NHEJ. This is especially relevant to cancers which are HR-defective as the incidence of unresolved DSBs may already be elevated. These may therefore become substrates for C-NHEJ if *POLQ* is also inhibited resulting in increased survival and genomic heterogeneity.

## Chapter 6 General Discussion

### 6.1 SUMMARY

This study had three broad aims that were dealt with in three separate results chapters:

- Chapter 3: To determine how the absence of POLQ impacts on the ability of cells to escape a telomere-driven crisis

This was achieved by forcing clonal populations of HAP1 and HCT116 WT and POLQ-deficient cell lines through telomere-driven crisis via infection with a DN-hTERT cassette. Crisis was defined by reduced growth kinetics, changes in cell morphology, including the presence of multi-nucleated cells, and the detection of telomere fusion events. Interestingly, both POLQ-deficient cell lines escaped crisis more rapidly than their associated WT cell lines. This was characterised by a more rapid return to pre-crisis growth rates for the HAP1 *POLQ*<sup>-DN-hTERT</sup> cell line. However, no reduction in growth rates, which is usually characteristic of crisis, was observed for clonal populations of HCT116 *POLQ*<sup>-DN-hTERT</sup> cells as they progressed through telomere-driven crisis.

- Chapter 4: To assess the role of POLQ in the NHEJ pathways utilised for telomere fusion

This was achieved by using telomere fusion PCR to detect telomere fusions for clonal populations of HAP1 and HCT116 WT and POLQ-deficient cell lines as they progressed through telomere-driven crisis. This analysis showed an increase in inter-chromosomal fusions for both POLQ-deficient cell lines compared with the WT cell lines. In contrast, a reduction in the abundance of 17p:17p intra-chromosomal fusion events was observed for the HCT116 *POLQ*<sup>-DN-hTERT</sup> cell line compared with the

HCT116  $WT^{DN-hTERT}$  cell line. Additionally, Transcription Activator-Like Effector Nuclease (TALEN) pairs were employed to induce DSBs in subtelomeric DNA in HAP1, HCT116 and RPE1<sup>hTERT</sup> WT and POLQ-deficient cell lines. The resulting telomere fusion events were detected using telomere fusion PCR. No significant differences in the abundance or type of fusion event were observed for the WT cell lines compared with the POLQ-deficient cell lines following nucleofection with 16p/21q TALEN pairs. However, an increase in the abundance of 17p:17p intra-chromosomal fusion events was observed for the HCT116 WT cell line compared with the HCT116  $POLQ^{-/-}$  cell line following nucleofection with 17p TALEN pairs.

- Chapter 5: To investigate how the loss of POLQ impacts on the evolving cancer genome

This was achieved by performing Illumina paired-end sequencing of fusion amplicons produced following 16p:21q dual-primer PCR for HAP1  $WT^{DN-hTERT}$  and HAP1  $POLQ^{-/-:DN-hTERT}$  clones as they progressed through telomere-driven crisis. A significantly greater proportion of genomic and inter-chromosomal telomere fusion events was identified for the HAP1  $POLQ^{-/-:DN-hTERT}$  clones compared with the HAP1  $WT^{DN-hTERT}$  clones. Greater variation in microhomology and insertion length distributions were observed for the HAP1  $WT^{DN-hTERT}$  cell line compared with the HAP1  $POLQ^{-/-:DN-hTERT}$  cell line. Additionally, whole genome sequencing was also performed on HAP1  $WT^{DN-hTERT}$  and HAP1  $POLQ^{-/-:DN-hTERT}$  clones before and after crisis. This analysis showed that a greater proportion of HAP1  $POLQ^{-/-:DN-hTERT}$  clonal populations had become polyclonal after crisis compared with the HAP1  $WT^{DN-hTERT}$  cell line.

## 6.2 GENERAL DISCUSSION

### 6.2.1 Telomere lengths in POLQ-deficient cell lines

Over 15 years ago, the chaos1 (chromosome aberration occurring spontaneously 1) mutation in mice that results in a major amino acid change from serine to proline in the *Polq1* gene was discovered (Shima et al. 2003). The uncharacterised gene product, DNA polymerase theta, had a C-terminal DNA polymerase and an N-terminal Hel308-type DNA helicase and was associated with defects in DNA repair or cell division (Shima et al. 2004). This discovery led to the purification of full-length human *POLQ* (Seki et al. 2003). Studies in *Drosophila melanogaster* later showed that the *POLQ* ortholog, *mus308*, operates independently of C-NHEJ and HR. In addition, cells with mutations in *mus308* and *spn-A*, a *RAD51* ortholog, were shown to be hyper-sensitive to ionizing radiation (Chan et al. 2010). Since then, research using mammalian cells has been performed to characterise the function of *POLQ* and has demonstrated an important role for *POLQ* in A-NHEJ-mediated repair of DSBs (Kent et al. 2015; Mateos-Gomez et al. 2015). It has been demonstrated that *POLQ* is involved in repairing DSBs induced by DNA damage throughout the genome. Furthermore, *POLQ*-mediated repair has also been observed at deprotected telomeres, although *POLQ* activity during telomere-driven crisis has not been explored. Therefore, the aim of this study was to investigate the role of *POLQ* in telomeric stability and telomere-driven crisis. As crisis is an important event in early tumorigenesis that may drive malignant progression, understanding the mechanisms that contribute to this event may lead to the identification of therapeutic targets.

Three distinct WT and *POLQ*-deficient human cell lines were used during this study: HAP1, RPE1<sup>hTERT</sup> and HCT116. STELA of telomere lengths in these cell lines revealed that *POLQ* disruption was associated with significantly longer XpYp telomere lengths for all three cell lines. This telomere length disparity was confirmed at other chromosome ends by a low resolution TRF assay which reveals the mean

telomere length based on all chromosome ends. This assay demonstrated that all three *POLQ*-deficient cell lines had longer mean telomere lengths than their WT counterparts. As these *POLQ*-deficient cell lines were not derived from the same parental populations as the WT cell lines, this discrepancy in telomere length could be a consequence of the CRISPR/Cas9 gene editing process employed to generate the *POLQ*-deficient cell lines. It is feasible that longer telomere lengths were necessary for cellular survival following the gene editing process. To test this hypothesis, CRISPR/Cas gene editing could be utilised to disrupt *POLQ* in clonal populations of the WT cell lines. This would allow for a direct comparison of telomere length before and after *POLQ* targeting. Alternatively, *POLQ*-deficient cells could be complemented with WT *POLQ* to see if this would restore shorter WT telomere lengths.

If *POLQ* disruption does affect telomere length, this would reveal a novel role for *POLQ* in telomere length maintenance. The three cell lines analysed were telomerase positive, therefore telomere elongation associated with the disruption of *POLQ* could be dependent on telomerase expression. To investigate this further, CRISPR/Cas gene editing to disrupt *POLQ* could be performed in clonal populations of somatic WT cells in which telomerase is not expressed alongside clonal populations of WT telomerase positive cells. Alternatively, this could be performed in fibroblasts alongside fibroblasts transfected with *hTERT* to investigate this mechanism in cells which are not pre-cancerous.

Aside from its polymerase activity, it has been shown that the helicase domain facilitates the interaction of *POLQ* with components such as Replication Protein A (RPA) (Mateos-Gomez et al. 2017) and RAD51 (Ceccaldi et al. 2015). It is therefore possible that interactions mediated by the helicase domain are involved in telomere maintenance mechanisms. It is currently understood that telomere length is regulated by the accessibility of telomeric DNA to telomerase, mediated by various components

of the shelterin complex. For example, POT1 binds to single-stranded 3'-overhang of TTAGGG repeats at the end of a telomere (Takai et al. 2011), limiting the access of telomerase to the terminal G residue resulting in the inhibition of telomere elongation (Lei et al. 2004).

It has been shown that knockdown of POT1 in human cells caused the accumulation of RPA at telomeric DNA, resulting in elongation of telomere length. Therefore, it is thought that POT1 antagonises the binding of RPA at telomeric ssDNA, but the mechanism involved remains unclear (Barrientos et al. 2008). When present at similar *in vitro* concentrations, RPA is able to outcompete POT1-TPP1 and bind as or more efficiently to telomeric ssDNA (Flynn et al. 2011). Furthermore, RPA is more abundant than POT1-TPP1 in primary and transformed human cells (Wold 1997; Takai et al. 2011). These data suggest that POT-TPP1 must associate with (an)other component(s) to successfully facilitate the removal of RPA from ssDNA.

Previously published data has already shown that the helicase domain of POLQ can facilitate the removal of RPA from ssDNA. It is therefore possible that POLQ is also able to facilitate the removal of RPA from telomeric ssDNA. In this situation, the disruption of POLQ may result in the accumulation of RPA at telomeric ssDNA, thereby preventing POT-TPP1 from binding to telomeric ssDNA. As the telomeric ssDNA is now accessible, telomere elongation could then be facilitated by telomerase.

To investigate this further, biochemical investigation could be undertaken to establish potential molecular interactions between POLQ, POT1 and telomeric DNA. For example, gel-shift assays could be repeated using POT1-TPP1 complexes and RPA to investigate the binding of both components to the telomeric ssDNA probe (Flynn et al. 2011). This could be repeated by co-incubating with POLQ to determine whether POLQ has any effect on telomeric ssDNA binding. Furthermore, pulldown assays using biotinylated telomeric ssDNA (Flynn et al. 2011) could be repeated in the

presence of POLQ to ascertain whether POLQ displaces RPA from ssDNA, allowing the binding of POT1 to ssDNA. This could be accompanied by experiments in which a myc-tag is added to POLQ. The addition of a myc-tag would facilitate the detection of POLQ to determine whether the protein localises to telomeric ssDNA.

### **6.2.2 Escape from telomere-driven crisis in HAP1 and HCT116 cell lines**

As telomerase is downregulated in the majority of normal human somatic cells, telomeres shorten following every round of cell division. Eventually, the loss of telomeric DNA results in insufficient chromosome end protection. This leads to the activation of the DNA damage response which triggers senescence or apoptosis. Telomere shortening is therefore considered to function as a tumour suppressor mechanism as it restricts the replicative lifespan of normal somatic cells which have accumulated mutation over time (Harley et al. 1990; Bodnar et al. 1998). However, if cells lack functional DNA damage cell cycle checkpoints, cells undergo telomere-driven crisis in which mounting telomere dysfunction becomes a source of genomic instability (Maciejowski and De Lange 2017). Genetic recombinations during crisis may result in the reactivation of telomerase which reconstitutes telomere function and restores proliferative capacity. The escape from crisis is therefore associated with a heavily rearranged, but stabilised genome, that may have acquired tumourigenic mutations. These extensive rearrangements are often observed in cancer genomes. Furthermore, reactivation of telomerase has been identified in ~90% of human cancers (Jafri et al. 2016).

Previously published data has demonstrated that telomere fusions can be mediated by C-NHEJ, A-NHEJ, SSA and BIR (Doksani 2019). Furthermore, C-NHEJ is associated with long-range inter-chromosomal telomere fusion, whereas A-NHEJ is associated with intra-chromosomal sister chromatid telomere fusion events that arise during telomere-driven crisis (Jones et al. 2014). It is suggested that the escape from crisis is influenced by the relative activities of DSB repair pathways which facilitate

the formation of inter- and intra- chromosomal telomere fusion events with telomeric and genomic loci. The Jones et al. (2014) study showed LIG3-deficient clonal populations of HCT116<sup>DN-hTERT</sup> failed to escape crisis and had increased levels of C-NHEJ-mediated inter-chromosomal fusion events.

By driving clonal populations of HCT116 WT and POLQ-deficient cells into telomere-driven crisis I was able to determine that POLQ is an important factor in facilitating the escape from crisis as the HCT116 POLQ-deficient cells escaped crisis more readily than the WT cell line. Strikingly, no change in growth kinetics or cell morphology were associated with telomere-driven crisis for the HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> clones. The HCT116 POLQ-deficient cell line also had longer starting telomere lengths and longer telomere lengths associated with the onset of crisis (as defined by the presence of telomere fusion events).

Driving clonal populations of HAP1 WT cells into telomere-driven crisis resulted in telomere erosion and the formation of telomere fusion events. However, the HAP1 POLQ-deficient cell line had extremely long starting telomere lengths (mean 18.74 kb) and whilst telomere erosion was observed, the shortest XpYp telomere lengths observed for the HAP1 *POLQ*<sup>-/-:DN-hTERT</sup> clones were between 2.87 and 8.55 kb (mean 5.14 kb). At these telomere lengths, I would expect shelterin components to bind to telomeric DNA, preventing telomere fusion. It is possible that telomere-driven crisis could have been initiated in this cell line by the accumulation of a few critically short telomeres at other chromosome ends (Counter et al. 1992) that were not studied. However, as telomeres are susceptible to DNA damage following replication stress and fork collapse (Gómez-Escoda and Wu 2018), telomere fusions observed for this cell line may have arisen due to replication stress (Liddiard et al. 2016). The increased abundance of genomic telomere fusion events for this cell line compared with the WT could also suggest that these fusions with non-telomeric loci potentially further transmit replication stress and drive the formation of further DSBs and fusion events.

As with the HCT116 cell line, the HAP1 *POLQ*<sup>:DN-hTERT</sup> cell line experienced less of a reduction in growth rates compared with the HAP1 WT<sup>DN-hTERT</sup> cell line, indicating that both cell lines had a greater tolerance for telomere fusion events.

For both HAP1 and HCT116 cell lines, telomere lengths were maintained following the detection of telomere fusion events and a return to pre-crisis growth rates. For all cell lines, the TRAP assay demonstrated that this was due to the reactivation of telomerase. As both HAP1 WT and *POLQ* cells line had become diploid early after transduction, it is likely that diploidisation was associated with long-term culturing of the cells. However, it is possible that this diploidisation provided cells with a survival advantage which contributed to the ability of these cells to escape crisis.

Whole genome sequencing of the HAP1 WT<sup>DN-hTERT</sup> and HAP1 *POLQ*<sup>:DN-hTERT</sup> cell lines before and after crisis revealed that a greater proportion of HAP1 *POLQ*<sup>:DN-hTERT</sup> clones were polyclonal after crisis compared with the HAP1 WT<sup>DN-hTERT</sup> clones. Changes in clonality may be a consequence of chromosomal fusion events that increase genomic heterogeneity and facilitate the escape from crisis.

### 6.2.3 The role of *POLQ* in telomere fusion events

I have shown that the disruption of *POLQ* in the HCT116 cancer cell line was associated with increased inter-chromosomal telomere fusion events and a decrease in 17p:17p sister chromatid fusion events. This was observed in two separate models, telomere-driven crisis and subtelomere-specific DSBs induced via nucleofection with TALEN plasmid pairs. This corroborates previously published data which shows that *POLQ* is an essential component of the A-NHEJ pathway (Wyatt et al. 2016) that is associated with intra-chromosomal telomere fusion events (Jones et al. 2014). Therefore, the disruption of *POLQ* results in the loss of the A-NHEJ pathway and a dependence on other forms of DSB repair such as C-NHEJ.

The increased abundance of C-NHEJ-mediated rearrangements (inter-chromosomal fusion events) observed for POLQ-deficient cell lines was confirmed by fusion amplicon sequencing data from HAP1<sup>DN-hTERT</sup> clones as they progressed through crisis. This data also revealed that an overall increase in the abundance of fusion events was observed for the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones compared with the HAP1 WT<sup>DN-hTERT</sup> clones. The presence of small regions of microhomology and small insertions at fusion junctions suggested that a common repair mechanism (C-NHEJ) (Chiruvella et al. 2013) was utilised by both cell lines with the exception of a few fusion events for the HAP1 WT<sup>DN-hTERT</sup> clones that had increased microhomology and insertion lengths that could be associated with A-NHEJ (Kent et al. 2015) or SSA. The increased proportion of C-NHEJ-mediated inter-chromosomal fusion events observed for the HAP1 *POLQ*<sup>:DN-hTERT</sup> clones compared with the HAP1 WT<sup>DN-hTERT</sup> clones suggests that the disruption of *POLQ* resulted in the upregulation of C-NHEJ-mediated repair of DSBs which facilitates inter-chromosomal telomere-fusion events. As this cell line escaped crisis more rapidly than the WT cell line, these data would suggest that C-NHEJ recombinations are more effective at enabling reactivation of telomere maintenance mechanisms. However, alterations at other genomic loci via these recombination events may also cause gross chromosomal rearrangements which have been detected in human malignancies (Lemée et al. 2010).

The increased proportion of inter-chromosomal fusion events in POLQ-deficient cells may be due to the incidence of DSBs which are elevated during telomere-driven crisis. It has been demonstrated that HR facilitates repair of ~50% of DSBs when the overall incidence of DSBs is low. However, increasing DSB frequency and genome instability causes suppression of repair pathways dependent on resection, such as HR and A-NHEJ, meaning cells are dependent on other resection-independent forms of repair such as C-NHEJ (Mladenov et al. 2019). Therefore, the increase in C-NHEJ-mediated inter-chromosomal telomere fusion events for the POLQ-deficient cell line

may be due to an elevated incidence of unresolved DSBs as A-NHEJ is inhibited which raises the overall DSB and instability to a threshold at which HR is suppressed. In addition, the threshold for HR/A-NHEJ may have already been surpassed in these cell lines as the incidence of DSB increased as a function of telomere shortening during telomere-driven crisis. It is therefore plausible that the HAP1 *POLQ<sup>:DN-hTERT</sup>* cell line is disproportionately dependent on C-NHEJ, causing an increased abundance of inter-chromosomal and genomic fusion events compared with the HAP1 *WT<sup>DN-hTERT</sup>* cell line.

In contrast to this, TALEN-induced DSBs did not result in the employment of different repair mechanisms in WT compared with *POLQ*-deficient cells. Presumably because these cells were not undergoing crisis before they underwent nucleofection with TALEN pairs, the low incidence of DSBs ensures that resection-dependent and independent mechanisms of DSB repair can be employed.

Whole genome sequencing data for HAP1 *WT<sup>DN-hTERT</sup>* and HAP1 *POLQ<sup>:DN-hTERT</sup>* clones showed that a greater proportion of HAP1 *POLQ<sup>:DN-hTERT</sup>* clones were polyclonal following crisis (75%) compared with HAP1 *WT<sup>DN-hTERT</sup>* clones (25%). The increased proportions of C-NHEJ-mediated inter-chromosomal fusion events for the HAP1 *POLQ<sup>:DN-hTERT</sup>* clones may have contributed to the increased polyclonality of the cell line due to the increased genomic heterogeneity conferred by these types of recombinations. This is reflected by the fusion data which revealed that the HAP1 *POLQ<sup>:DN-hTERT</sup>* clones had a significantly greater proportion of genomic fusion events that could be expected to promote widespread instability. This discrepancy in the proportion of polyclonal clones between the cell lines could suggest that genomic fusion events are more likely to result in recombinations that drive polyclonality and facilitate the upregulation of telomerase, thus allowing the *POLQ*-deficient cell lines to escape crisis more readily than the WT cell lines. This increase in polyclonality is also consistent with the growth curves that showed that the *POLQ*-deficient cell lines

rapidly transited crisis, indicating that multiple single cells were capable of independently re-establishing/upregulating telomerase activity.

As it has previously been shown that the POLQ helicase domain is able to actively dissociate RPA from ss-DNA, it is possible that disruption of *POLQ* allows increased RPA-ssDNA binding (Mateos-Gomez et al. 2017). Increased RPA binding may result in an increased tolerance to ssDNA, preventing fork collapse and thus suppressing BIR, thereby promoting cell survival and reducing replication stress (Malkova and Ira 2013). This may explain why the *POLQ*-deficient cell lines experienced less of a reduction in growth rates as they progressed through telomere-driven crisis. Previously published data has demonstrated that *POLQ* is able to actively displace both RAD51 (Ceccaldi et al. 2015) and RPA (Mateos-Gomez et al. 2017) from ssDNA in an ATP-dependent manner. As RPA and RAD51 are required for HR, SSA and BIR (Scully et al. 2019), it may be that disruption of *POLQ* results in the inhibition of these repair pathways.

A greater proportion of inter-chromosomal fusion events were observed for both *LIG3*-deficient (Jones et al. 2014) and *POLQ*-deficient cell lines (*Chapters 4&5*) compared with WT controls. Whilst it is likely that these are mediated by C-NHEJ, it is possible that other repair pathways such as SSA and BIR contribute to the fusion of critically short telomeres (Doksani 2019). However, whilst the *POLQ*-deficient cells escaped more readily than the WT cells, the *LIG3*-deficient cells failed to escape crisis.

A-NHEJ is suppressed in these *LIG3*-deficient cells, however, as *POLQ* is expressed, it may be that the *POLQ* helicase domain is able to suppress HR, SSA and BIR via its interactions with RPA or RAD51. As *LIG3* is also required for BER (Krokan and Bjørås 2013) these *LIG3*-deficient cells are restricted in terms of DNA repair that they can employ. Whilst C-NHEJ is possible, if resection by MRN-CtIP occurs, repair is dependent on either HR or A-NHEJ and cannot become substrates for C-NHEJ (Shamanna et al. 2016). Whilst this repair may be resolved by *LIG1* during S/G<sub>2</sub>-

phases of the cell cycle (Liddiard et al. 2018), if the incidence of DSBs are already such that HR and A-NHEJ are suppressed (i.e. in a crisis context of heightened instability and DNA damage), and if SSA, BIR and BER are also inhibited, these DSBs may remain unresolved. The accumulation of unresolved DSBs could trigger cell death and therefore render LIG3-deficient cells incapable of escaping crisis (Jones et al. 2014). To test this hypothesis, HCT116 cells deficient in both LIG3 and POLQ could be generated and forced into telomere-driven crisis to determine whether the disruption of *POLQ* could help facilitate the escape from crisis in these cells.

Alternatively, it may be that POLQ is an important factor for influencing which repair pathway is utilised for DSB repair. The disruption of POLQ caused a significant increase in the abundance of C-NHEJ-mediated inter-chromosomal fusion events compared with WT cells. It could be possible that in the absence of POLQ, C-NHEJ is employed. However, in LIG3-deficient cells, POLQ promotes A-NHEJ which cannot be ligated by LIG3. As POLQ inhibits HR, SSA and BIR, these resected DSBs remain unresolved and thus result in cell death in the absence of POLQ during crisis.

### **6.3 LIMITATIONS, CONCLUSIONS AND FUTURE DIRECTIONS**

#### **6.3.1 Limitations**

There are limitations associated with this study that should be considered. For example, it is possible that CRISPR/Cas9 targeting of *POLQ* resulted in a truncated variant of the protein that retains some function. The *POLQ*-targeted mutations could be confirmed by DNA sequencing. However, as there are currently no effective antibodies for detecting POLQ protein expression in the cells used in this study could not be confirmed.

### 6.3.2 Conclusions

This study corroborates previous data that shows POLQ is an essential component of the A-NHEJ pathway of DSB repair that is associated with intra-chromosomal telomere fusion events during telomere-driven crisis.

This study further suggests that disruption of POLQ, and therefore A-NHEJ, causes an increase in C-NHEJ-mediated inter-chromosomal telomere fusion events and a clear reduction in A-NHEJ mediated intra-chromosomal events. As POLQ-deficient cell lines escaped crisis more readily than their WT counterparts, these data suggest that C-NHEJ-mediated inter-chromosomal fusion events result in recombinations that result in the reactivation of telomere maintenance mechanisms.

These findings have implications for whether POLQ represents a valid therapeutic target. The overexpression of POLQ has been identified in a number of human cancers and has been linked with poor clinical outcome for breast cancer patients with HR-deficiency. A synthetic lethal approach has been suggested which would combine chemotherapy and POLQ inhibition for the treatment of HR-deficient tumours. It is thought that POLQ would increase the effectiveness of chemotherapy for initiating cell death as chemotherapy-induced DSBs are not subject to repair by either HR or A-NHEJ. However, data presented here could suggest that the inhibition of POLQ in HR-deficient cells could result in increased levels of C-NHEJ. Furthermore, in these HR-deficient cells, the high incidence of DSBs initiated by chemotherapy may result in the suppression of A-NHEJ, further facilitating the use of C-NHEJ to repair DSBs. As this study has shown that the disruption of *POLQ* resulted in a more rapid escape from crisis, potentially due to increased C-NHEJ-mediated recombination, the upregulation of C-NHEJ in HR-deficient cancer cells could confer increased tumoural heterogeneity and more rapid clonal evolution.

### 6.3.3 Future directions

Much more research is required to comprehend the diverse roles of POLQ in human cells.

For example, further investigation into why POLQ-deficient cells have longer telomeres is required to ascertain whether POLQ is involved in telomere maintenance mechanisms.

As both HAP1 and HCT116 POLQ-deficient cell lines escaped telomere-driven crisis more rapidly than their WT counterparts, more research is necessary to understand how the disruption of POLQ facilitates crisis escape. A greater proportion of HAP1 *POLQ*<sup>-DN-hTERT</sup> clones become polyclonal during crisis and an increase in the abundance of inter-chromosomal telomere fusion events was observed compared with the HAP1 WT<sup>DN-hTERT</sup> clones. Whether these factors contributed to a more timely escape needs to be addressed. To do so, paired-end sequencing of fusion amplicons detected following 17p:21q dual-primer PCR for HCT116 WT<sup>DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cells transiting crisis could be performed. This would allow characterisation of inter- compared with intra-chromosomal fusion events that occur during telomere-driven crisis for the two cell lines. In addition, whole genome sequencing of HCT116 WT<sup>DN-hTERT</sup> and HCT116 *POLQ*<sup>-/-:DN-hTERT</sup> cells before and after crisis, alongside parental lines could be performed. This could enable the analysis of the clonality of the two cell lines to see if the observations made for the HAP1<sup>DN-hTERT</sup> cell lines are also observed for the HCT116<sup>DN-hTERT</sup> cell lines.

The increased proportion of inter-chromosomal telomere fusion events observed for the POLQ-deficient cell lines are thought to be mediated via C-NHEJ. To confirm that C-NHEJ-mediated repair is necessary for the telomere fusion events that arose in the POLQ-deficient cell lines that progressed through crisis, clonal populations of HCT116 *POLQ*<sup>-/-:LIG4</sup><sup>-/-</sup> could be forced into telomere-driven crisis. In addition, as these cells are C-NHEJ- and A-NHEJ-deficient, this analysis could inform whether

other forms of DSB repair aside from NHEJ are able to precipitate recombinations that mediate the escape from crisis.

The POLQ helicase domain has been shown to actively displace both RPA and RAD51 from ssDNA. It is possible that the in HCT116<sup>DN-hTERT</sup> LIG3-deficient cells, the suppression of A-NHEJ coupled with the downregulation of other DNA repair mechanisms such as BIR, HR and SSA (which require RPA or RAD51) via the activity of the POLQ helicase, resulted in the accumulation of unresolved DSBs that ultimately caused cell death and a failure to escape crisis. To test this hypothesis, HCT116 cells that are deficient in both LIG3 and POLQ could be generated and forced into telomere-driven crisis.

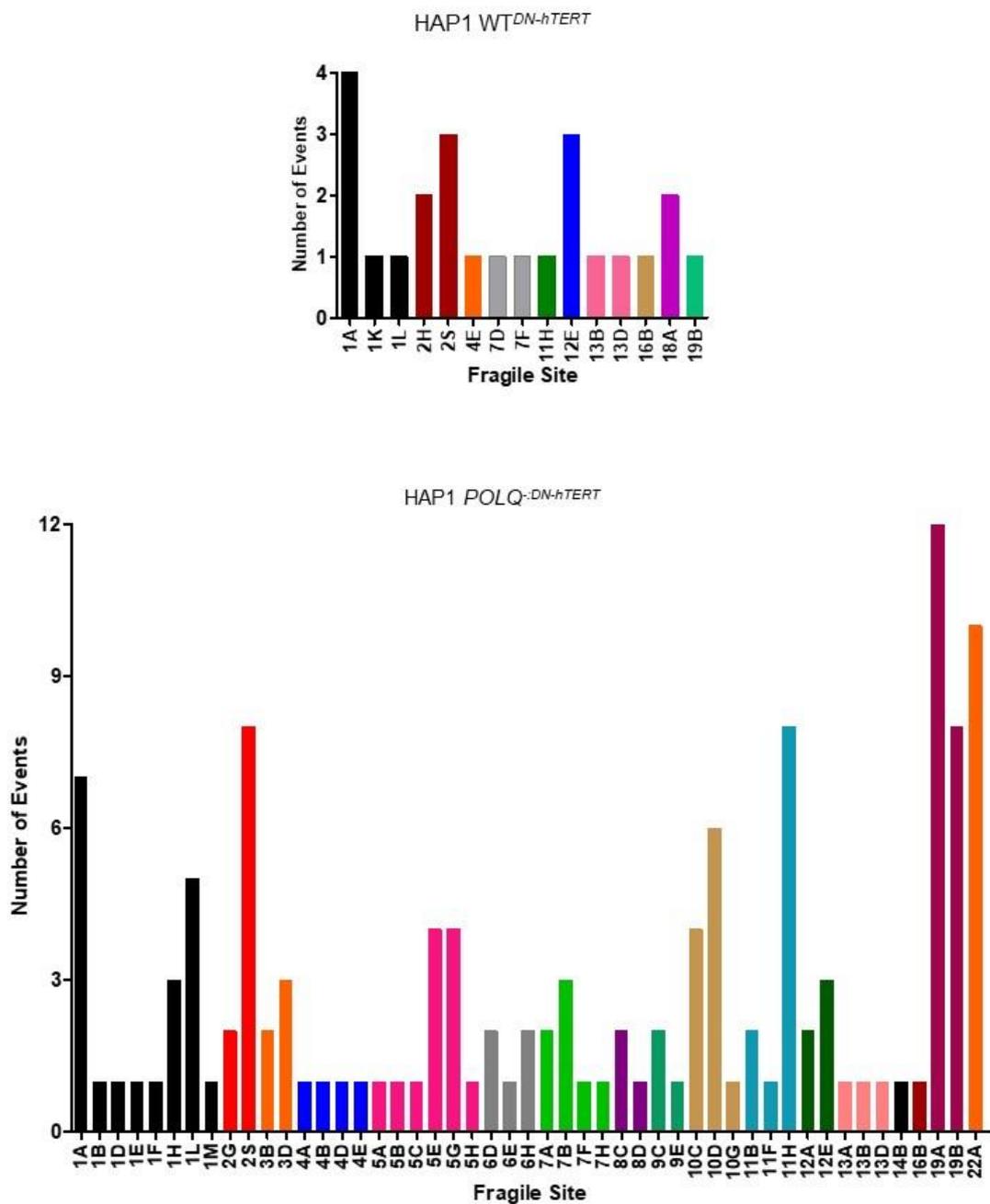
POLQ over-expression has been observed in several human cancers and is linked with adverse clinical outcome. In addition, it has been demonstrated that HR-deficient tumours are dependent on POLQ for survival. As such, it is postulated that the targeted inhibition of POLQ in HR-deficient cancers (such as EOCs) will improve the efficacy of chemotherapy. However, data presented here suggests that POLQ inhibition may lead to increased genomic instability which could aid malignant progression. To test the impact of POLQ inhibition on HR-deficiency, CRISPR/Cas disruption of *POLQ* in HR-deficient cancer cells could be performed. These cells could be subject to DSBs using CRISPR/Cas or TALEN pairs to analyse the impact of these DSBs on cellular survival and genome stability. However, it would be important to culture these cells without inducing DSBs to determine whether the inhibition of POLQ mediates cell survival and malignant progression.

Altogether, much more research is required to enhance our understanding of the many roles that this intriguing enzyme has in both healthy and cancerous human cells.

## Appendix

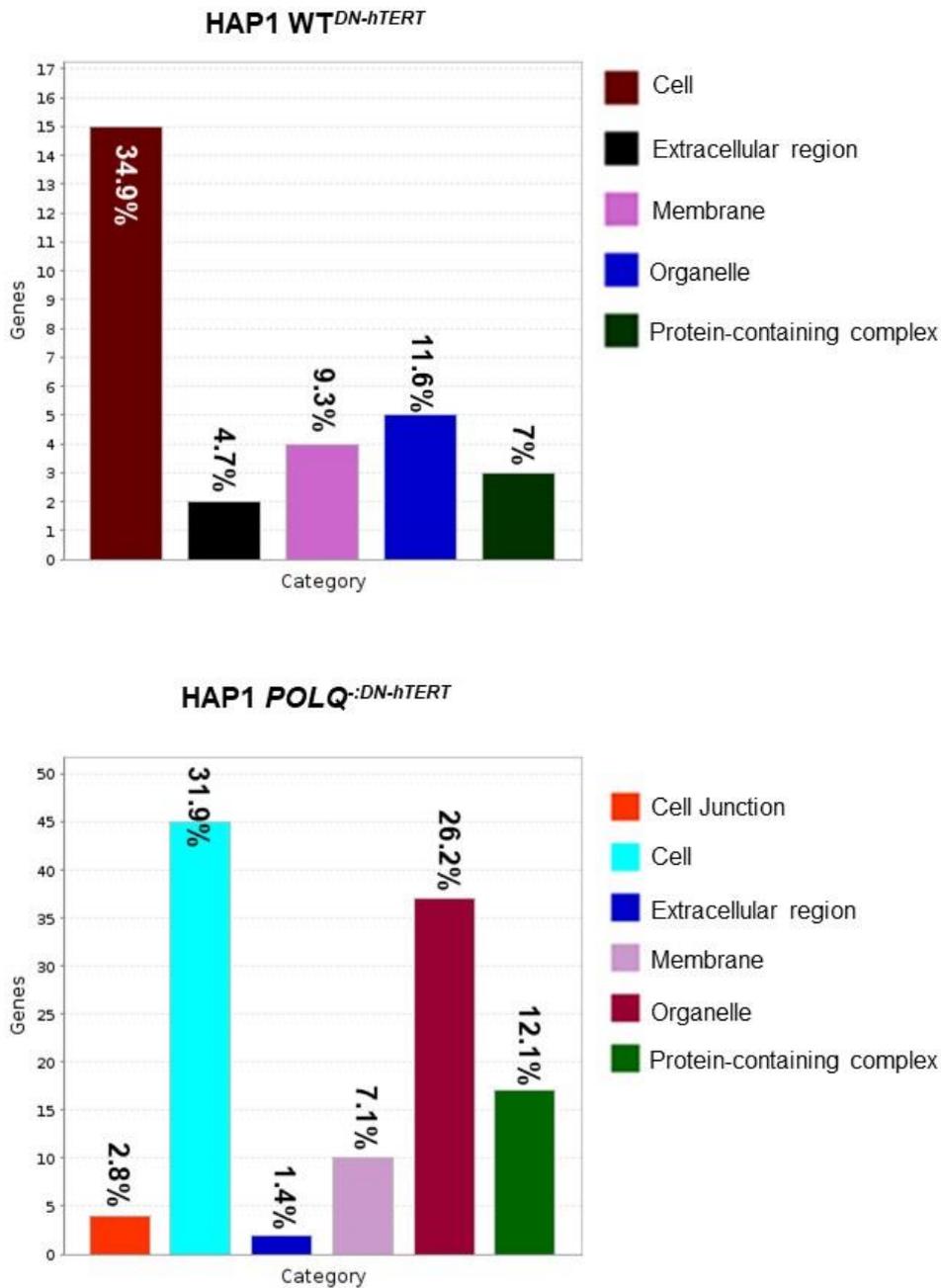
Application	Primer name	Primer sequence (5' to 3')
<b>STELA</b>	5p5	GGAGCAGCATTCTCTTCACCACAG
	9p2	CACATTCCTCATGTGCTT ACG
	11q13B	CAGACCTTGGAGGCACGGCCTTCG
	16prev1	CACTTATTAGTTCCAGTCTCTG
	17pseq1rev	GAATCCACGGATTGCTTTGTGTAC
	XpYpC	CAGGGACCGGGACAAATAGAC
	XpYpE2	TTGTCTCAGGGTCCTAGTG
	Teltail	TGCTCCGTGCATCTGGCATC
	Tel2	TGCTCCGTGCATCTGGCATCTAACCCCT
<b>FUSION</b>	17p6	GGCTGAACTATAGCCTCTGC
	16p1	TGGACTTCTCACTTCTAGGGCAG
	21q1	CTTGGTGTCGAGAGAGGTAG
<b>GENOMIC SEQUENCING</b>	<i>POLQ</i> FORWARD EXON 1	AGTAGAAGCCCAATGGGGTATG
	<i>POLQ</i> REVERSE EXON 1	GAGGTTTGAGTTTGAAGACTGGC
	<i>POLQ</i> FORWARD EXON 2	TCGTTTGAGGCTTTTAAGGAATTT
	<i>POLQ</i> REVERSE EXON 2	GCACTGCCTCATTACCTTT
	<i>LIG4</i> FORWARD	ATGATTGCATATTTTGTGTTGAAGCC
	<i>LIG4</i> REVERSE	GCAGCTAGCATTGGTTTAAATGC

**Supplementary Table 1:** Table listing all primers used during this study.



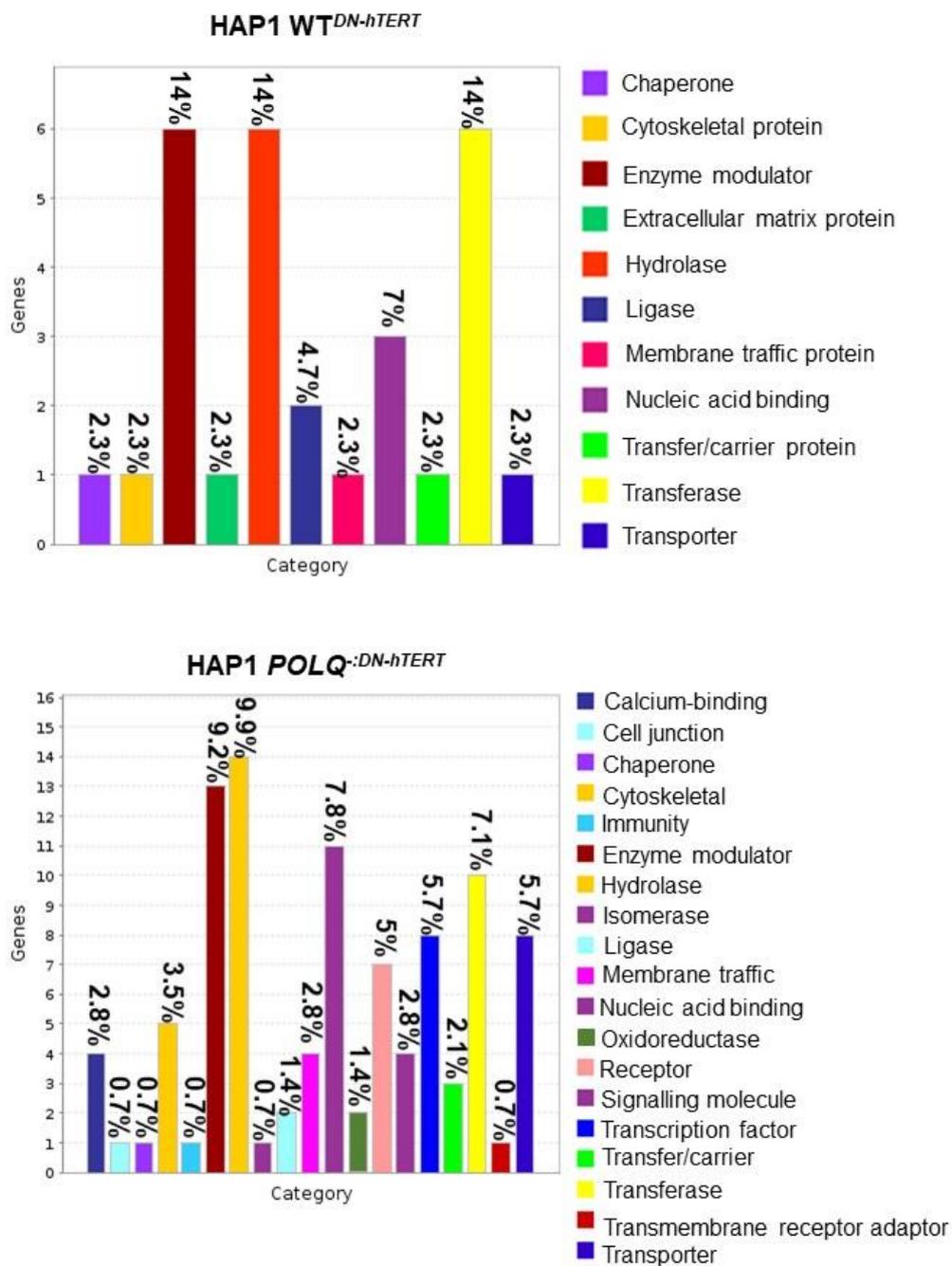
**Supplementary Figure 1:** Bar chart showing the total number of genomic telomere fusion events that overlapped with fragile sites for the HAP1 WT<sup>DN-hTERT</sup> (top) and the HAP1 POLQ<sup>DN-hTERT</sup> clones (bottom).

## Cellular Component

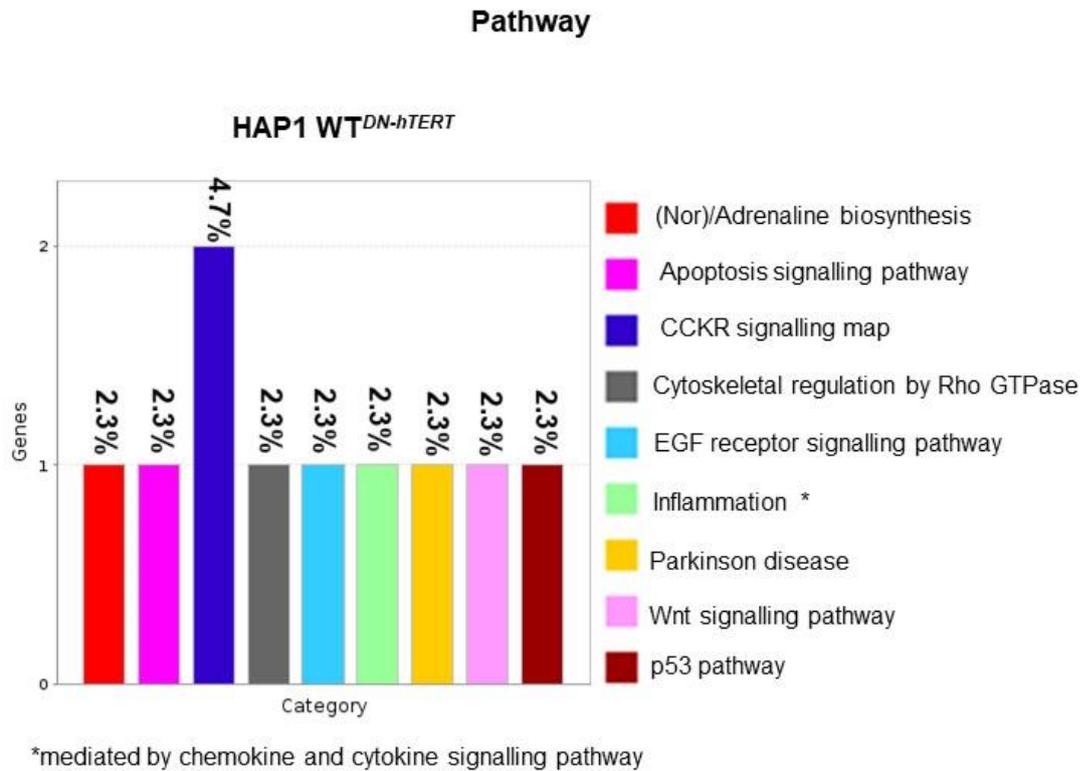


**Supplementary Figure 2:** Bar charts showing the proportion of genomic telomere fusion events that overlapped with genes associated with different cellular functions for the HAP1 WT<sup>DN-hTERT</sup> (top) and the HAP1 POLQ<sup>DN-hTERT</sup> cell line (bottom).

## Protein Class



**Supplementary Figure 3:** Bar charts showing the proportion of genomic telomere fusion events that overlapped with genes associated with different classes of proteins for the HAP1 WT<sup>DN-hTERT</sup> (top) and the HAP1 POLQ<sup>:DN-hTERT</sup> cell line (bottom).



**HAP1 POLQ<sup>-/-</sup>DN-hTERT**

Pathways that represent 1.4% of the total number of genes	
Pathway	
	Axon guidance
	Ionotropic glutamate receptor
	Integrin signalling pathway
	Insulin/IGF pathway-protein kinase B signalling
	FGF signalling pathway
	Endothelin signalling
	PDGF signalling pathway

**Supplementary Figure 4:** Bar chart showing and tables listing the proportion of genomic telomere fusion events that overlapped with genes associated with different molecular pathways for the HAP1 WT<sup>DN-hTERT</sup> (top bar chart) and the HAP1 POLQ<sup>-/-</sup>DN-hTERT cell line (bottom tables).

**HAP1 POLQ<sup>-/-</sup>DN-hTERT**

<b>Pathways that represent 0.7% of the total number of genes</b>
<b>Pathway</b>
Apoptosis signalling
5HT2 type receptor mediated signalling
Alpha adrenergic receptor signalling
Inflammation mediated by chemokine and cytokine signalling
Pentose phosphate pathway
Endogenous cannabinoid signalling
Hedgehog signalling
Glycolysis
Transcription regulation by bZIP transcription factor
Thyrotropin-releasing hormone receptor signalling
T cell activation
Oxytocin receptor mediated signalling
2-arachidonoylglycerol biosynthesis
p38 MAPK pathway
Parkinson disease
PI3 kinase pathway
Circadian clock system
Oxidative stress response
Histamine H1 receptor mediated signalling
Blood coagulation
Angiotensin II-stimulated signalling through G proteins and beta-arrestin
Metabotropic glutamate receptor group I pathway

**HAP1 POLQ<sup>-/-</sup>DN-hTERT**

<b>Pathway</b>	<b>Percentage</b>
Metabotropic glutamate receptor group III	2.1%
Angiogenesis	2.1%
Hypoxia response via HIF activation	2.1%
p53 pathway	2.1%
VEGF signalling	2.1%
Cadherin signaling pathway	2.1%
Gonadotropin-releasing hormone receptor pathway	2.8%
Huntington disease	2.8%
CCKR signaling map	3.5%
Heterotrimeric G-protein signalling	3.5%
Wnt signalling	5.0%

**Supplementary Tables:** Tables listing the Dysgu ID, Feature (F) (microhomology = H, blunt = B and insertions = I) of the fusion junction, the fusion junction feature sequence (M in uppercase, l in lowercase) any genes that overlapped with the fusion junction, the fusion junction location coordinates and any fragile sites (FS) or CpG islands that overlapped with the fusion junction (N = no overlap). The clone is listed at the top of the table in bold. (KO refers to *POLQ* clones).

WT C1							
ID	Feature	Sequence	Gene	Location start	Location end	FS	CPG
4	microh	ATGAGGTCAG	TBCK	chr 4:106200620	chr4:106200695		N
5	microh	CTTCTCCAGC		chr X: 81776853	chr X: 81666972		N
6	blunt			chr 15:69698293	chr15:69698370		N
7	microh	A		chr 8: 89668920	chr 8:89668998		N
8	microh	TCC	SCTR	chr2:119463125	chr2:119463228	FRA2S	N
10	microh	T		chr 13:114353033	chr13:114353095		N
11	microh	G		chr 12:76238526	chr 12:76328606		N
13	microh	T	DMD	chr X:31287130	chrX:31287215		N
19	microh	CTCCCAAAGTGCTGGGATTACAG		chr 9:2224506	chr 9:2224579		N
30	microh	TGACTTCC	WASH2P	chr2:113597285	chr2:113597395	FRA2S	N
31	microh	CTCTGC		chr2:113604129	chr 11:113604207	FRA2S	N
32	microh	TG		chr13:61755468	chr13:61755539	FRA13B	N

WT C11							
ID	Feature	Sequence	Gene	Location start	Location end	FS	CPG
1238	microh	CTGAGGCAGGAGAATGGCGTGAACCC	CCSER1	chr4:91423280	chr4:91423354	FRA4E	N
1241	microh	GCCTGCGCCA	WASHC1	chr9:14579	chr9:14686		N

WT C7							
ID	feature	sequence	Gene	Location start	Location end	FS	CPG
364	blunt		SLC12A2	chr5:128085829	chr5:128085923		N
365	blunt		EIF2AK2	chr2:37133243	chr2:37133345		N
366	microh	CTTTGCT		chr8:64312370	chr8:6431245		N
367	microh	CTCCTGTCAC	KIAA1328	chr18:37172978	chr18:37173065	FRA18A	N
372	microh	AACAAATGACA	IARS	chr20:31356465	chr20:31356514		N
373	microh	T		chr9:92211128	chr9:92211193		N
374	microh	GGCGCG	MLIP	chr6:54240881	chr6:54240935		N
375	blunt			chr2:187065086	chr2:187065181	FRA2H	N
376	microh	GCCCAG		chr20:41383267	chr20:41383364		N
377	microh	TGCAGTGGCACAATCTTGGCTCAATGCAA	SMARCA4	chr19:10990246	chr19:10990359	FRA19B	N
378	microh	TAGT	TPCN1	chr12:113296117	chr12:113296152	FRA12E	N
380	microh	AACCC		chr17:52236648	chr17:52236714		N
381	microh	AGCAGTGAAG		chr12:114172179	chr12:114172277	FRA12E	N
383	microh	C		chr9:97886132	chr9:97886198		N
384	None			chr4:161287351	chr4:161287470		N
386	microh	T		chr7:12890702	chr7:12890751		N
392	blunt		ADAMTSL3	chr15:83933489	chr15:83933561		N
395	None		ROCK1	chr18:21076567	chr18:21076611		N
396	microh	CTGTGTCATT		chr8:208099	chr8:208163		N
397	microh		SPATA13	chr13:2404443	chr13:24044510		N

WT C8							
ID	F	Sequence	Gene	Location start	Location end	FS	CPG
796	M	GGCGGG	PITPNC1	chr17:67524167	chr17:67524216		N
797	M	CT	DIRC3	chr2:217648781	chr2:217648873		N
805	M	CAGAACTGTG		chr3:196151404	chr3:196151494		N
806	M	CCTCCCaAAGTGCTGGGAT		chr16:11229511	chr16:11229629		N
808	M	TT	MOB2	chr11:1476430	chr11:1476535		N
810	M	CAAA		chr8:114153457	chr8:114153538		N
811	M	TGGT		chr12:130240458	chr12:130240504	FRA12E	N
812	M	CTCAA	RNF14	chr5:141968447	chr5:141968564		N
813	M	ACAGGCG	ZBTB17	chr1:15970593	chr1:15970694	FRA1A	N
814	M	GGCTCACACCTATAATCCCAACACTTTGGGAGGCC	IGF2BP3	chr7:23376597	chr7:23376669		N
816	M	AGA	THBS3	chr1:155207690	chr1:155207819		N
818	B			chr1:188863709	chr1:188863806	FRA1K	N
821	B		TIE1	chr1:43311798	chr1:43311917		N
822	M	AGCATCCCCAG		chr1:16618023	chr1:16618088	FRA1A	N
827	M	CT		chr12:65604044	chr12:65604144		N
828	M	CT	IDNK	chr9:83624402	chr9:83624493		N
829	M	GCCTCCCAAAAGTGCTGGGATTACAGGCCTGAGCC	DIEXF	chr1:209853611	chr1:209853720		N
833	B		EFCAB12	chr3:129417735	chr3:129417826		N
836	M	GCGCC		chr14:48359151	chr14:48359225		N
837	M	ATCATG	DNAJC6	chr1:65302776	chr1:65302881	FRA1L	N
840	M	CTGCTGGC		chr8:206612	chr8:206664		N
843	B		SPPL2A	chr15:50756341	chr15:50756420		N

WT C12							
ID	F	Sequence	Gene	Location start	Location end	FS	CPG
1555	M	AG	PODXL2	chr3:127653251	chr:127653376		N
1556	M	C		chr5:12766445	chr5:12766606		N
1558	B		SPTBN5	chr15:41852226	chr15:41852289		N
1559	M	TT		chr11:102494632	chr11:102494693		N
1560	B		CD82	chr11:44578413	chr11:44578472		N
1561	M	CCCGTCTCTAC	HS6ST2	chrX:132790973	chrX:132791044		N
1562	M	CACTGGG		chr17:47013765	chr17:47013827		N
1563	M	GCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGC	LCAT	chr16:67941295	chr16:67941392	16B	N
1565	M	AG	STX12	chr1:27808085	chr1:27808206	1A	N
1566	M	GATTT	DAB2IP	chr9:121768431	chr9:121768539		N
1567	M	TCT	CSRNP3	Chr2:165505163	chr2:165505201		N
1568	B		RBFOX1	chr16:5529159	chr16:5529201		N
1570	M	CATTT		chr22:19440257	chr22:19440367		N
1571	B			chr2:187259118	chr2:187259240	2H	N
1573	M	C	MINOS1	chr1:19614236	chr1:19614302	1A	N
1574	I	acaaa		chr4:134356697	chr4:134356803		N
1576	M	GCAG	MIRLET7BHG	chr22:46111280	chr22:46111351		N
1579	I	cggc		chr8:206449	chr8:206538		N
1580	M	ACCCCTGTCTCTAC		chr13:99144007	chr13:99144067	13D	N
1582	M	AGCTCAAAAC		chr1:203291515	chr1:203291598		N
1583	B		UQCC1	chr20:35361188	chr20:35361243		N
1587	I	gtcaagggtcacgggtcagcgttaaggggtcaaggtc		50807780	50807853		N
1590	B			chr3:83232316	chr3:83232427		N
1592	B		KIAA1328	chr18:37010471	chr18:37010450	18A	N
1593	B		DNAL1	chr14:73677968	chr14:54493728		N
1595	B			chr7:44921113	chr7:44921209	7D	N
1596	M	G	RELT	chr11:73388444	chr11:73388532	H	N
1597	B			chr14:98028040	chr14:98028155		N
1600	M	CA		chrX:24367058	chrX:24367167		N
1602	I	acactgaccaggagaaaaaccaaccaatagaaacaggcccagac	ZBED4	chr22:49879382	chr22:49879474		N

KO C2						
ID	F	Sequence	Gene	Location start	Location end	FS CPG
2078	M	TGTGTTTC	FRMPDF	chrX:12617759	12617850	
2079	M	GGCAGAGAAAGACA	ERCC6L2	chr9:95947370	95947479	
2080	M	CAGCCTCAGGT	SOX21-AS1	chr13:94763885	94763995	
2083	M	CT	LINC02641	chr10:123453054	123453151	10D
2084	M	A	BLM	chr15:90814291	90814336	
2085	M	G	PPM1B	chr2:44216419	44216507	2S
2086	M	C		chr20:51300196	51300307	
2087	M	CTT		chr17:27448140	27448251	
2088	M	AA	PLS3	chrX:115614873	115614990	
2090	M	AAAA	TRIOBP	chr22:37715521	37715627	22A
2091	B			chr11:59289363	59289441	
2092	M	C		chr13:47595098	47595158	
2093	M	C	ZC3H7B	chr22:41332485	41332569	22A
2094	M	CC	MAX	chr14:65065985	65066048	14B
2095	B			chr17:27448076	27448163	
2096	B		SLCO2A1	chr3:133932250	133932329	
2097	B		SPHK2, DBP	chr19:48629997	48630105	19A
2098	M	A	CDH10	chr5:24590562	24590682	5E
2099	M	CCC		chr3:42447766	42447892	
2100	B			chr2:102553490	102553591	2S
2103	M	TC		chr13:95415823	95415882	13D
2104	I	gg	MTA1	chr14:105450110	105450194	
2105	B		PART1	chr5:60547667	60547763	5H
2106	B		CDH10	chr5:24590562	24590651	5E
2107	M	A	SRSF4	chr1:29157395	29157492	
2108	B		PLS3	chr3:133932250	133932329	
2111	M	GC	GRIK5	chr19:42054187	42054262	19A
2113	M	GCTG		chr1:39794642	39794718	
2115	B			chr7:128861948	128862062	
2117	M	C		chr14:97690905	97690974	
2118	M	GT	EPC1	chr10:32345575	32345670	
2119	M	G		chr11:109467433	109467520	
2120	M	C	SPHK2	chr19:48630184	48630254	19A 116

2121	M	CT		chr16:30589363	30589425		
2122	B		XXYL1	chr3:195170488	195170569		
2123	B		CDK14	chr7:91064350	91064445		
2124	M	**	CLPS	chr6:35796664	35796767	6H	
2126	B		UBR5	chr8:102286329	102286402	8A	
2127	I	ctt	WDR82	chr3:52256920	52257034		
2128	M	G	CCAR1	chr10:68771891	68771932	10C	
2130	B		MIR5096	chr22:37947982	37948096	22A	
2131	M	T	NKAIN2	chr6:124450734	124450820		
2132	M	C		chr1:229074027	229074078	1H	
2134	M	TCC		chr6:148056043	148056138		
2135	M	GCCCTTCC		chr16:3646928	3647055		
2136	B			chr2:113606305	113606397	2S	
2138	M	GT	SRSF4	chr1:29157548	29157654		
2141	B		SLC2A9	chr4:9852969	9853089	4A	
2142	M	TGCCACCAC		chr19:7308635	7308750	19B	
2143	M	ACTTTGGGAGGCC	ASTN2	chr9:117275822	117275863	9E	
2144	B		PDE1A	chr2:182262371	182262508	2G	
2145	M	GCCACTGCA	SCARA5	chr8:27931926	27932039		
2146	B			chr2:113604344	113604449	2S	
2147	B		CHKA	chr11:68077697	68077726	11H	
2149	B		OGFOD3	chr17:82413076	82413200		
2150	B			chr3:88566102	88566206		
2151	I	aaagtgct	TRPM3	chr9:70786232	70786297		
2152	M	A	MIR5096	chr22:37715763	37715887	22A	
2153	I	c		chr4:62726448	62726562		
2154	B		CLTCL1	chr22:19217338	19217479		
2155	M	TTTT		chr12:37344898	37344982		
2156	M	AGC	F2	chr11:46730011	46730087		
2157	M	G		chr20:51300180	51300243		
2158	M	GA	ATXN7	chr3:63952250	63952315		
2159	M	G	WDR82	chr3:52256893	52257004		
2160	M	TCAT	WVOX	chr16:78999249	78999381		
2161	M	G		chr19:35977121	35977189	19A	

\*\*GGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGCGGATCACGAGGTCAGGAG

ID	F	Sequence	Gene	Location S	End	FS	CGP
2162	M	TGGAT		chr9:120332458	120332512		
2163	I	a	NKAIN2	chr6:124450670	124450719		
2164	M	TGTG	LIMCH1	chr4:41635629	41635668		
2168	B		CDK5RAP2	chr9:120446297	120446401		
2169	B			chr3:32460135	32460220		
2170	B		SCPEP1	chr17:57004401	57004436		
2172	M	C	GSE1	chr16:85653857	85653913		
2177	B		OGFOD3	chr17:82413076	82413200		
2178	B			chr3:162742041	162742142		
2181	B			chr4:96128771	96128812	4E	
2182	M	G	CHKA	chr11:68077945	68077981	11H	
2183	B		DLGAP1	chr18:4355868	4355930		
2184	B			chr6:148056202	148056248		
2185	B		ZBTB8OS	chr1:32641500	32641640		
2186	M	TG	GRHL2	chr8:101561485	101561599		
2188	M	CAGGAG		chr5:155302655	155302778		
2189	M	AT	SUPT3H	chr6:45081600	45081649	6H	
2190	B		EEF1A1	chr6:73518862	73518953	6D	
2192	M	TCTG		chr2:176867983	176868020	2G	
2193	M	GCC	EFTUD1P1	chr15:84089653	84089717		
2194	I	ttaaattgggtgggattccagct		chr12:124685114	124685196	12E	
2196	M	GCAG		chr12:37344850	37344945		
2199	B	aaaaaaca	MIR5096	chr22:37947897	37947970	22A	
2201	B			chr14:82079150	82079241		
2203	B			chr1:39794546	39794630		
2204	B			chr13:114352995	114353087		
2205	M	CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCAC		chr11:22392033	22392135		
2207	B			chr11:109467592	109467680		
2209	B		EEF1A1	chr6:73518832	73518953	6D	

KO C3							
ID	F	Sequence	Gene	Location start	Location end	FS	CPG
3111	M	CC		chr2:61562697	61562816	2S	
3112	M	TCCTTT	FCFR10P	chr6:167036571	167036656		
3113	M	GGT	NUP50	chr22:45185803	45185923	221	
3114	M	GACAGAG		chr3:63054233	63054286	3B	
3115	M	GCTGG	ERC2	chr3:56400155	56400247		
3117	M	AAAGGGGA		chr4:65793787	65793889		
3119	B		TSC22D2	chr3:150414508	150414602	3D	
3120	M	C	LPAR3	chr1:84857245	84857279	1L	
3122	M	CT	AGO4	chr1:35821799	35821827		
3123	B			chr7:127345277	127345369		
3124	M	C	SETD2	chr3:47149449	47149542		
3125	ins	att		chr15:56591902	56592004		
3126	M	CCT	PPP2R2D	chr10:131903363	131903416	10D	
3128	M	G	PIK3R1	chr5:68288310	68288425		26
3129	M	C	KCTD19	chr16:67304952	67305005	16B	
3130	B		NBEA	chr13:35357959	35358078	13A	
3132	B		GFRAL	chr6:55356148	55356266		
3133	M	CGAGACCATCCTGG	TM2D1	chr1:61711375	61711460	1L	
3134	M	A		chr11:118881045	118881160	11B	
3135	M	GGAGGCTGAGGCAG	JAKMIP2	chr5:147709744	147709841		
3136	B		DIRC3-AS1	chr2:217315968	217316049		
3137	M	TCATC		chr1:93921801	93921858	1D	
3138	M	GGGACGGCA	MYT1L	chr2:2130171	2130190		
3139	M	GTGG		chr6:28429002	28429118		
3141	M	GGCTGG	C11orf84	chr11:63826064	63826151	11H	
3142	B		IL1RAPL2	chrX:105478310	105478374		
3144	M	TCTT	SLC39A5	chr12:56236159	56236262		
3145	M	GC		chr2:30363263	30363304		
3147	B		PLIN3	chr19:4845889	5760117	19B	
3148	B		TSC2	chr16:2080974	2081014		

ID	F	Sequence	Gene	Location start	Location end	FS	CPG
3149	M	GGATGGGCGCGGTGGCTCAC	RERE	chr1:8531136	8531192	1A	
3150	M	CG		chr2:30363712	30363817		
3151	ins	g		chr15:99047782	99047833		
3153	M	GCCAAGGTGGGC		chr14:97687463	97687530		
3154	ins	taa		chr12:54158306	54158350	12A	
3155	M	TT	PIK3R1	chr5:68288249	68288313		
3156	M	TTCTGCTCAGTTC	SH2B3	chr12:111432345	111432439	12E	
3159	M	GCT	MAN1A1	chr6:119182288	119182383		
3160	B		PLCB1	chr20:8615596	8615701		
3161	I	caatcaaaataggttcaactctgtgagttgaatgcacacatcacaagaagct	APOL2	chr22:36230207	36230261	22A	
3162	B		C4orf47	chr4:185440759	185440843		
3163	B		POLR2A	chr17:7497770	7497819		
3164	M	T	C11orf84	chr11:63825977	63826049	11H	
3166	M	C		chrX:46029608	46029683		
3167	M	TG	NDRG3	chr20:36729306	36729339		
3169	B		LPAR3	chr1:84857432	84857505	1L	
3170	B		PLXNA4	chr7:132418617	105478374	7H	
3172	M	CAC	UQCC1	chr20:35361186	35361243		
3174	M	CC	NDRG3	chr20:36729306	36729393		
3176	M	T		chr14:34896780	34896879		
3177	M	AAA	C4orf47	chr4:185440759	185440843		
3180	M	CTCGGCCTCCC	EDARADD	chr1:236434352	236434396	1H	
3181	B		TSC22D2	chr3:150414511	150414615	3D	

KO C5							
ID	F	Sequence	Gene	Location start	Location end	FS	CPG
3912	B			chr13:46802382	46802484		
3914	B			chr18:53736460	53736558		
3915	B		SLC25A4C	chr19:19081916	19081974	FRA19B	
3916	M	CTTCTCTC	SHROOM3	chr4:76629619	76629723		
3917	I	tataatgattatatt		chr18:44259937	44260020		
3918	M	TTG	PICALM	chr11:85963949	85964043	FRA11F	
3919	B			chr9:129393579	129393672		
3920	B			chr11:71523416	71523484	FRA11H	
3921	M	GAAGTCCAA		chr11:71523334	71523437	FRA11H	
3922	M	T		chr13:91356345	91356430		
3924	M	TA		chr12:37921780	37921849		
3931	B		ANK3	chr10:60206600	60206689	FRA10C	
3932	B		RNF14	chr5:141976449	141976556		
3933	B		URI1	chr19:29938255	29938355		
3934	M	A		chr6:153206315	153206391		
3936	B		BICD1	chr12:32248422	32248516		
3937	B			chr4:25505714	25505754	FRA4D	132
3938	B		EEF2K	chr16:22216822	22216867		
3939	M	AGGTG	C5orf66	chr5:135228132	135228166	FRA5C	
3940	M	C		chr19:53300521	53300669	FRA19A	
3941	B			chr2:153072357	153072473	FRA2S	
3942	M	TGA		chr4:139230648	139230716		
3944	M	G		chr17:76992065	76992099		
3945	B			chr4:179685078	179685188		
3947	M	TT	ATE1-AS1	chr10:121932839	121932940	FRA10D	
3949	M	A	BAIAP2L1	chr7:98329312	98329343	FRA7F	
3950	M	TT	GPC3	chrX:133565792	133565854		

3951	B			chr19:50142376	50142436	FRA19A	
3952	M	A		chr2:36094111	36094158		
3953	M	TT	RBM20	chr10:110765000	110765062	FRA10D	
3957	M	AAG		chr16:23013267	23013299		
3961	M	CTA		chr6:107834504	107834587	FRA6F	
3962	I	aaaagaaaacaaaacaaaaaggcactcca	COPG1	chr3:129264230	129264316		
3963	B			chr4:137458827	137458930		
3965	B			chr14:98949374	98949488		
3968	B			chr18:44259926	44259962		
3970	B		AGO4	chr1:35821799	35821888		
3972	M	GAGG	FGF12	chr3:192620148	192620242		
3973	B		NMNAT1	chr1:9964867	9964958	FRA1A	
3974	M	T		chr20:34190579	34190635		
3977	M	CCTGG		chr4:25505904	25506003	FRA4D	CpG_132
3978	B			chr12:37921543	37921641		
3983	M	GCC	PCNX1	chr14:70970098	70970175		
3984	B			chr3:31070096	31070214		
3985	M	TGTG	FAT3	chr11:92366402	92366449		
3989	I	cat		chr3:182419533	182419653		
3992	M	AGGAT	PTK2	chr8:140743468	140743521	FRA8D	
3994	M	GGATGC		chr1:39616698	39616805		
4000	B			chr12:62970603	62970708		
4001	I	tctttgat		chr1:20044678	20044713	FRA1A	
4006	M	TTCA		chr3:36323439	36323531		
4008	I	ctgtc	BICD1	chr12:32248317	32248362		
4009	M	AG		chr4:54652019	54652108	FRA4B	
4014	B			chr6:153206315	153206391		
4018	M	GCGC		chr22:50806970	50807055		
4019	B			chr10:44353242	44353353	FRA10G	
4020	B			chr5:66131017	66131140		
4027	M	AGG		chr5:179796880	179796913	FRA5G	
4029	M	TG	SLIT3	chr5:168676709	168676778	FRA5G	
4030	M	AT		chr7:7009262	7009422	FRA7B	
4031	B			chr2:117164633	117164696	FRA2S	
4034	B			chr19:50142324	50142436	FRA19A	

KO C8							
ID	feature	sequence	Gene	Location start	Location end	Fragile site	CPG
4911	M	G		chr20:63216498	63216589		60
4913	B			chr6:125315693	125315829		
4914	I	a		chr11:43949307	43949359		
4917	I	ct	LRBA	chr4:150964351	150964462		
4919	B		CSNK1D	chr17:82248344	82248376		
4921	B			chr1:62399287	62399328	1L	
4922	I	ag	CSNK1D	chr17:82248509	82248574		
4923	M	TCCA	ATF3	chr1:212571047	212571120		
4924	M	TG	SLC35A3	chr1:99983171	99983279	1E	
4926	M	G	DPEP1	chr16:89635820	89635896		
4927	M	AA	TMEM54	chr1:32898223	32898253		
4928	M	TGC	LINCO1476	chr17:59467458	59467580		
4929	M	GC	SPECC1L	chr22:24406091	24406201		
4930	M	C		chr7:55279076	55279147	7A	
4931	I	g		chr6:125315663	125315772		
4932	M	T	RBM23	chr14:22908461	22908567		
4933	M	C		chr1:155657939	155658013		
4934	I	caagagcaac		chr15:67991284	67991328		
4935	I	ggaa	RBM23	chr14:22908523	22908659		
4936	M	TG	CADPS	chr3:62811286	62811379	3B	
4938	M	T	TSC22D1	chr13:44552546	44552587		
4939	M	AC		chr14:27150830	27150853		
4942	M	TTTGACCT	WWOX	chr16:78857913	78858011		
4944	M	T		chr1:59163319	59163401	1B	
4945	M	G		chr2:81783576	81783660	2S	
4947	I	c		chr1:1898862	1898951	1A	
4948	M	AG	NR_130776	chr5:96044397	96044492	5B	
4949	M	CA		chr1:112091095	112091199		
4950	M	A		chr15:70248855	70248930		
4951	M	C		chr22:29391342	29391404		
4954	B		PACRG	chr6:163115726	163115856	6E	
4955	B		FAM13B	chr5:138000710	138000777		
4956	M	G		chr1:62399154	62399273	1L	

4957	B			chr17:81962151	81962262		
4959	I	tgactctgtccag		chr17:28411392	28411424		
4961	B		HNRNPM	chr19:8469149	8469225	19B	
4962	M	G		chr1:24279847	24279887	1A	
4964	M	G	TNPO3	chr7:128957898	128958007		
4965	M	T		chr13:59017845	59017917	13B	
4967	B		KIAA0355	chr19:34325408	34325514	19A	
4969	B			chr8:119359693	119359779	8C	
4971	M	TGG	HNRNPM	chr19:8469028	8469080	19B	
4972	M	AG		chr20:37246074	37246115		
4973	B		TACR2	chr10:69414454	69414540	10C	
4975	M	A	KIAA0355	chr19:34325408	34325507	19A	
4976	I	g		chr1:116884598	116884618		
4977	M	TGGT		chr19:16239254	16239302	19B	
4979	I	gcgt	CAPN8	chr1:223547872	223547984		
4980	M	CCT	LINC02149	chr5:15205568	15205635		
4981	M	TAGTT	FAM13B	chr5:138000612	138000650		
4982	M	CTGTGTCATT		chr19:248306	248370	19B	
4987	B		MYO1H	chr12:109442719	109442789	12E	
4989	M	T		chrX:71280916	71280945		
4990	M	GCCACCGCACC		chr2:232594710	232594828		
4991	M	GG		chr1:24279650	24279695	1A	
4993	M	TGT	ZC3H7B	chr22:41317993	41318027	22A	
4994	B			chr1:1898851	1898887	1A	
4996	M	CCAGGCTGGTCT		chr11:59628840	59628915		
4997	M	GCC	INO80	chr15:41088315	41088359		
4998	B			chr11:43949307	43949394		
4999	I	t		chr7:41192721	41192782		
5000	I	g	GPC5	chr13:92543674	92543737		
5001	M	T	SH3PXD2A	chr10:103738036	103738067	10D	
5004	M	GA	ADAM5	chr8:39383753	39383849		
5005	B		CDH18	chr5:19663035	19663109	5E	
5011	M	A	EEF2	chr19:3982960	3982989	19B	29
5012	B			chr9:29627706	29627824	9C	
5013	B		SCAPER	chr15:76382509	76382559		

ID	feature	sequence	Gene	Location start	Location end	Fragile site	CPG
5014	M	CTTTGGCTTAAGAA		chr2:38277460	38277509		
5015	B		NLK	chr17:28062074	28062151		
5017	M	AC		chr20:7973593	7973650		
5018	B		DDN-AS1	chr12:49002289	49002394	12A	
5019	I	ccg		chr11:63794608	63794637	11H	
5023	M	CAGGTCCAG		chr14:20572144	20572225		
5024	B			chr20:55002965	55003062		
5025	M	AC	GPC5	chr13:92543541	92543612		
5026	B		TSC22D1	chr13:44552399	44552505		
5027	M	AG		chr18:64926908	64926945		
5035	M	C	SLC19A3	chr2:227690473	227690517		
5039	M	TGA	ADGRL3	chr4:61893533	61893577		
5042	B		AFF2	chrX:148590001	148590110		
5045	B		MED12L	chr3:151360946	151360986	3D	
5049	B			chr10:17265383	17265493		
5058	M	TA	PRDM5	chr4:120810161	120810207		
5061	B		NR_110668	chr1:224222499	224222628	1H	
5064	B			chr7:55279085	55279182	7A	
5065	B		MTAP	chr9:21850583	21850618	9C	
5066	M		AFF2	chrX:148589863	148589944		
5067	M	AG		chr14:27150945	27150976		
5068	M	A		chr20:63216483	63216513		
5072	M	AGG	LINC02149	chr5:15205568	15205635		60
5075	M	C	NR_110817	chr3:35109793	35109894		
5078	B			chr11:63794521	63794595	11H	
5083	B		ATXN10	chr22:45747031	45747120	22A	
5088	B			chr13:91091777	91091843		
5089	M	GAAAAGA		chr14:44186753	44186839		

KO C10							
ID	F	Sequence	Gene	Location start	Location end	FS	CPG
6341	M	T	STIM1	chr11:4046702	4046778		
6344	M	C	FAF2	chr5:176456451	176456519	5G	
6346	B		GRIK4	chr11:120752745	120752834	11B	
6347	B		PGGT1B	chr5:115215860	115215990		
6350	M	CCCA		chr17:8377087	8377087		
6351	M	GAC	NUDT17	chr1:145845703	145845784	1F	40
6353	M	GC		chr2:31322310	31322389		
6355	B		DLC1	chr8:13599738	13599821		
6358	M	G	FBXO46	chr19:45730943	45731047	19A	35
6360	M	A		chr20:46317755	46317808		
6361	M	G		chr6:141427595	141427652		
6362	M	A		chr5:19158743	19158836	5E	
6363	M	G		chr6:170607365	170607421		
6366	M	G		chr14:38116766	38116814		
6368	M	CA		chr22:39310701	39310737	22A	
6369	M	TTTA	GRM8	chr7:127214966	127215007		
6372	I	ac	TRIM71	chr3:32848853	32848982		
6373	I	ct	TMEM110-MUSTN1	chr3:52893006	52893043		
6374	M	A	CASC6	chr6:91641668	91641705	6G	
6376	M	A		chr6:84900082	84900148		
6379	M	TC		chr5:31067112	31067153		
6380	M	AGC	SUPT20HL2	chrX:24312708	24312755		
6381	B		ASH1L	chr1:155391524	155391618		
6382	B		UPK1A	chr19:35669093	35669139	19A	
6383	M	A	TMPO	chr12:98520903	98521000		
6385	M	CAG		chr7:3254498	3254543	7B	
6389	M	CA		chrX:24367046	24367167		
6391	M	T	UPK1A	chr19:35669241	35669293	19A	
6394	I	ttttttgta		chr7:3254624	3254687	7B	
6396	M	CCATCCGTG		chr11:786590	786663		
6397	M	C	FAM222B	chr17:28771642	28771703		
6400	M	CTG	LCOR	chr10:96860817	96860946	10D	

6401	M	CAGG	MIR548N	chr1:99693468	99693524	1M	
6405	M	TC		chr12:43133896	43133936		
6406	M	TT	ACR	chr22:50742483	50742600		
6408	B			chr12:107918716	107918763		
6420	I	ct	HNRNPUL2-BSCL2	chr11:62720787	62720853		
6421	B			chr17:68010806	68010922		
6429	M			chr2:31322310	31322357		
6430	M	CC	PDZD2	chr5:32103565	32103564	5A	
6433	B		FAF2	chr5:176456403	176456519	5G	
6436	B			chr3:193822091	193822142		
6438	B	C	ZNF423	chr16:49587691	49587737		

## References

Aguirrezabalaga, I. et al. 1995. THE CROSS-LINKING AGENT HEXAMETHYLPHOSPHORAMIDE PREDOMINANTLY INDUCES INTRA-LOCUS AND MULTILOCUS DELETIONS IN POSTMEIOTIC GERM-CELLS OF DROSOPHILA. *Genetics* 139(2), pp. 649-658.

Alder, J. K. et al. 2018. Diagnostic utility of telomere length testing in a hospital-based setting. *Proceedings of the National Academy of Sciences of the United States of America* 115(10), pp. E2358-E2365. doi: 10.1073/pnas.1720427115

Allshire, R. C. et al. 1989. HUMAN TELOMERES CONTAIN AT LEAST 3 TYPES OF G-RICH REPEAT DISTRIBUTED NON-RANDOMLY. *Nucleic Acids Research* 17(12), pp. 4611-4627. doi: 10.1093/nar/17.12.4611

Arana, M. E. et al. 2008. Low-fidelity DNA synthesis by human DNA polymerase theta. *Nucleic acids research* 36(11), p. 3847. doi: 10.1093/nar/gkn310

Arnoult, N. and Karlseder, J. 2015. Complex interactions between the DNA-damage response and mammalian telomeres. *Nature Structural & Molecular Biology* 22(11), p. 859. doi: 10.1038/nsmb.3092

Arnoult, N. et al. 2010. Replication Timing of Human Telomeres Is Chromosome Arm-Specific, Influenced by Subtelomeric Structures and Connected to Nuclear Localization (Replication Timing of Human Telomeres). *PLoS Genetics* 6(4), p. e1000920. doi: 10.1371/journal.pgen.1000920

Artandi, S. E. and Attardi, L. D. 2005. Pathways connecting telomeres and p53 in senescence, apoptosis, and cancer. *Biochemical and Biophysical Research Communications* 331(3), pp. 881-890. doi: 10.1016/j.bbrc.2005.03.211

Aubert, G. et al. 2012. Telomere length measurement—Caveats and a critical assessment of the available technologies and tools. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 730(1-2), pp. 59-67. doi: 10.1016/j.mrfmmm.2011.04.003

Audebert, M. et al. 2004. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *The Journal of biological chemistry* 279(53), p. 55117. doi: 10.1074/jbc.M404524200

Aviv, A. et al. 2011. Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR. *Nucleic acids research* 39(20), p. e134. doi: 10.1093/nar/gkr634

Baird, D. M. et al. 2000. High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: Implications for telomere biology and human evolution. *American Journal Of Human Genetics* 66(1), pp. 235-250.

Baird, D. M. et al. 1995. Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *The EMBO journal* 14(21), p. 5433. doi: 10.1002/j.1460-2075.1995.tb00227.x

Baird, D. M. et al. 2003. Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nature Genetics* 33(2), pp. 203-207. doi: 10.1038/ng1084

Balajee, A. S. et al. 1994. ANALYSIS OF RESTRICTION ENZYME-INDUCED CHROMOSOME-ABERRATIONS IN THE INTERSTITIAL TELOMERIC REPEAT SEQUENCES OF CHO AND CHE CELLS BY FISH. *Mutation Research* 307(1), pp. 307-313. doi: 10.1016/0027-5107(94)90304-2

Baldwin, A. S. 1996. The NF-kappa B and I kappa B proteins: new discoveries and insights. *Annual review of immunology* 14, p. 649.

Bankevich, A. et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* 19(5), pp. 455-477.

Barrientos, K. S. et al. 2008. Distinct Functions of POT1 at Telomeres. *Molecular and Cellular Biology* 28(17), p. 5251. doi: 10.1128/MCB.00048-08

Batté, A. et al. 2017. Recombination at subtelomeres is regulated by physical distance, double-strand break resection and chromatin status. *EMBO Journal* 36(17), pp. 2609-2625. doi: 10.15252/emj.201796631

Baur, J. A. et al. 2001. Telomere position effect in human cells. *Science (New York, N.Y.)* 292(5524), p. 2075. doi: 10.1126/science.1062329

Ben-David, U. and Amon, A. 2020. Context is everything: aneuploidy in cancer. *Nature reviews. Genetics* 21(1), p. 44. doi: 10.1038/s41576-019-0171-x

Bhargava, R. et al. 2016. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends in Genetics* 32(9), pp. 566-575. doi: 10.1016/j.tig.2016.06.007

Bianchi, A. et al. 1997. TRF1 is a dimer and bends telomeric DNA. *EMBO Journal* 16(7), pp. 1785-1794. doi: 10.1093/emboj/16.7.1785

Bignell, G. R. et al. 2010. Signatures of mutation and selection in the cancer genome. *Nature* 463(7283), p. 893. doi: 10.1038/nature08768

- Blackburn, E. H. and Collins, K. 2011. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harbor perspectives in biology* 3(5), doi: 10.1101/cshperspect.a003558
- Bodnar, A. G. et al. 1998. Extension of life-span by introduction of telomerase into normal human cells. *Science (New York, N.Y.)* 279(5349), p. 349. doi: 10.1126/science.279.5349.349
- Bothmer, A. et al. 2017. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nature Communications* 8, doi: 10.1038/ncomms13905
- Boulton, S. J. and Jackson, S. P. 1996. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *The EMBO journal* 15(18), p. 5093. doi: 10.1002/j.1460-2075.1996.tb00890.x
- Britt-Compton, B. and Baird, D. M. 2006. Intra-allelic mutation at human telomeres. *Biochem Soc Trans* 34(Pt 4), pp. 581-582. doi: 10.1042/bst0340581
- Britt-Compton, B. et al. 2009. Short telomeres are preferentially elongated by telomerase in human cells. *FEBS Letters* 583(18), pp. 3076-3080. doi: 10.1016/j.febslet.2009.08.029
- Brown, W. R. A. et al. 1990. Structure and polymorphism of human telomere-associated DNA. *Cell* 63(1), pp. 119-132. doi: 10.1016/0092-8674(90)90293-N
- Bryan, T. M. et al. 1997. The telomere lengthening mechanism in telomerase-negative immortal human cells does not involve the telomerase RNA subunit. *Human molecular genetics* 6(6), p. 921. doi: 10.1093/hmg/6.6.921
- Cai, Y. and Sablina, A. A. 2016. Cancer-associated chromosomal deletions: Size makes a difference. *Cell Cycle* 15(21), pp. 2850-2851. doi: 10.1080/15548627.2016.1204869
- Cannan, W. J. and Pederson, D. S. 2016. Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin. *Journal of Cellular Physiology* 231(1), pp. 3-14. doi: 10.1002/jcp.25048
- Capper, R. et al. 2007. The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes & Development* 21(19), pp. 2495-2508. doi: 10.1101/gad.439107
- Cawthon, R. M. 2002. Telomere measurement by quantitative PCR. *Nucleic acids research* 30(10), p. e47. doi: 10.1093/nar/30.10.e47

- Cawthon, R. M. 2009. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic acids research* 37(3), p. e21. doi: 10.1093/nar/gkn1027
- Ceccaldi, R. et al. 2015. Homologous-recombination-deficient tumours are dependent on Pol $\theta$ -mediated repair. *Nature* 518(7538), p. 258. doi: 10.1038/nature14184
- Celli, G., B. and De Lange, T. 2005. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature Cell Biology* 7(7), p. 712. doi: 10.1038/ncb1275
- Cesare, A. J. and Griffith, J. D. 2004. Telomeric DNA in ALT Cells Is Characterized by Free Telomeric Circles and Heterogeneous t-Loops. *Molecular and Cellular Biology* 24(22), p. 9948. doi: 10.1128/MCB.24.22.9948-9957.2004
- Cesare, A. J. and Reddel, R. R. 2010. Alternative lengthening of telomeres: models, mechanisms and implications. *Nature Reviews Genetics* 11(5), p. 319. doi: 10.1038/nrg2763
- Chan, S. H. et al. 2010. Dual Roles for DNA Polymerase Theta in Alternative End-Joining Repair of Double-Strand Breaks in Drosophila. *Plos Genetics* 6(7), doi: 10.1371/journal.pgen.1001005
- Chang, H., H. Y. et al. 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology* 18(8), doi: 10.1038/nrm.2017.48
- Chang, H. H. Y. et al. 2016. Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. *The Journal of biological chemistry* 291(47), p. 24377. doi: 10.1074/jbc.M116.752329
- Chaudhuri, A. R. and Nussenzweig, A. 2017. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nature Reviews Molecular Cell Biology* 18(10), doi: 10.1038/nrm.2017.53
- Chen, L.-Y. and Lingner, J. 2013. CST for the grand finale of telomere replication. *Nucleus* 4(4), pp. 277-282. doi: 10.4161/nucl.25701
- Chiruvella, K. K. et al. 2013. Repair of double-strand breaks by end joining. *Cold Spring Harbor perspectives in biology* 5(5), p. a012757. doi: 10.1101/cshperspect.a012757
- Cleal, K. et al. 2019. Chromothripsis during telomere crisis is independent of NHEJ, and consistent with a replicative origin. *Genome research* 29(5), p. 737. doi: 10.1101/gr.240705.118

- Counter, C. M. et al. 1992. Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *The EMBO journal* 11(5), p. 1921. doi: 10.1002/j.1460-2075.1992.tb05245.x
- Deng, Y. et al. 2008. Telomere dysfunction and tumour suppression: the senescence connection. *Nature Reviews Cancer* 8(6), p. 450. doi: 10.1038/nrc2393
- Deng, Z. et al. 2009. TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Molecular Cell* 35(4), pp. 403-413. doi: 10.1016/j.molcel.2009.06.025
- Deniz, M. et al. 2017. Increased single-strand annealing rather than non-homologous end-joining predicts hereditary ovarian carcinoma. *Oncotarget* 8(58), pp. 98660-98676. doi: 10.18632/oncotarget.21720
- Dilley, R. L. and Greenberg, R. A. 2015. ALternative telomere maintenance and cancer. *Trends in cancer* 1(2), pp. 145-156.
- Doheny, J. G. et al. 2008. Telomeric Position Effect—A Third Silencing Mechanism in Eukaryotes (Telomeric Position Effect). *PLoS ONE* 3(12), p. e3864. doi: 10.1371/journal.pone.0003864
- Doksani, Y. 2019. The Response to DNA Damage at Telomeric Repeats and Its Consequences for Telomere Function. *Genes* 10(4), doi: 10.3390/genes10040318
- Doksani, Y. and de Lange, T. 2014. The role of double-strand break repair pathways at functional and dysfunctional telomeres. *Cold Spring Harbor Perspectives in Biology* 6(12), doi: 10.1101/cshperspect.a016576
- Donaldson, A. D. and Nieduszynski, C. A. 2019. Genome-wide analysis of DNA replication timing in single cells: Yes! We're all individuals. *Genome Biology* 20(1), doi: 10.1186/s13059-019-1719-y
- Ducray, C. et al. 1999. Telomere dynamics, end-to-end fusions and telomerase activation during the human fibroblast immortalization process. *Oncogene* 18(29), p. 4211. doi: 10.1038/sj.onc.1202797
- Durkin, S. G. and Glover, T. W. 2007. Chromosome fragile sites. *Annual review of genetics* 41, p. 169.
- Dwight, T. et al. 2018. structural rearrangements in metastatic pheochromocytomas. *Endocrine-related cancer* 25(1), p. 1. doi: 10.1530/ERC-17-0306
- Eisfeldt, J. et al. 2017. , an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000Research* 6, p. 664. doi: 10.12688/f1000research.11168.2

- Escribano-Díaz, C. et al. 2013. A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice. *Molecular Cell* 49(5), pp. 872-883. doi: 10.1016/j.molcel.2013.01.001
- Fagagna, F. D. A. D. et al. 2003. A DNA damage checkpoint response in telomere-initiated senescence. *Nature* 426(6963), p. 194. doi: 10.1038/nature02118
- Faridounnia, M. et al. 2018. Function and Interactions of ERCC1-XPF in DNA Damage Response. *Molecules* 23(12), doi: 10.3390/molecules23123205
- Feinberg, A. P. and Vogelstein, B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical biochemistry* 132(1), pp. 6-13.
- Fernandez-Vidal, A. et al. 2014. A role for DNA polymerase  $\theta$  in the timing of DNA replication. *Nature communications* 5(1), pp. 1-10.
- Flynn, R. L. et al. 2011. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* 471(7339), p. 532. doi: 10.1038/nature09772
- Fouquerel, E. et al. 2019. Targeted and Persistent 8-Oxoguanine Base Damage at Telomeres Promotes Telomere Loss and Crisis. *Molecular Cell* 75(1), pp. 117-130.e116. doi: 10.1016/j.molcel.2019.04.024
- Garcia-Diaz, M. and Bebenek, K. 2007. Multiple Functions of DNA Polymerases. *Critical Reviews in Plant Sciences* 26(2), pp. 105-122. doi: 10.1080/07352680701252817
- Gisselsson, D. et al. 2000. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proceedings of the National Academy of Sciences of the United States of America* 97(10), p. 5357. doi: 10.1073/pnas.090013497
- Greider, C. W. 1999. Telomeres Do D-Loop–T-Loop. *Cell* 97(4), pp. 419-422. doi: 10.1016/S0092-8674(00)80750-3
- Greider, C. W. and Blackburn, E. H. 1985. Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell* 43(2), pp. 405-413. doi: 10.1016/0092-8674(85)90170-9
- Griffith, J. D. et al. 1999. Mammalian telomeres end in a large duplex loop. *Cell* 97(4), p. 503. doi: 10.1016/S0092-8674(00)80760-6
- Guilliam, T. A. et al. 2015. Primase-polymerases are a functionally diverse superfamily of replication and repair enzymes. *Nucleic Acids Research* 43(14), pp. 6651-6664. doi: 10.1093/nar/gkv625

- Guo, H. et al. 2017. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Research* 27(2), pp. 165-183. doi: 10.1038/cr.2016.128
- Gómez-Escoda, B. and Wu, P.-Y. J. 2018. The organization of genome duplication is a critical determinant of the landscape of genome maintenance. *Genome research* 28(8), p. 1179. doi: 10.1101/gr.224527.117
- Hakem, R. 2008. DNA-damage repair; the good, the bad, and the ugly. *EMBO Journal* 27(4), pp. 589-605. doi: 10.1038/emboj.2008.15
- Halvorsen, T. L. et al. 1999. Telomerase Activity Is Sufficient To Allow Transformed Cells To Escape from Crisis. *Molecular and Cellular Biology* 19(3), p. 1864. doi: 10.1128/MCB.19.3.1864
- Harley, C. et al. 1990. TELOMERES SHORTEN DURING AGING OF HUMAN FIBROBLASTS. *Nature* 345(6274), pp. 458-460. doi: 10.1038/345458a0
- Hastie, N. D. and Allshire, R. C. 1989. Human telomeres: fusion and interstitial sites. *Trends in Genetics* 5(C), pp. 326-331. doi: 10.1016/0168-9525(89)90137-6
- Hastings, P. J. et al. 2009. Mechanisms of change in gene copy number. *Nature Reviews Genetics* 10(8), p. 551. doi: 10.1038/nrg2593
- Hasty, P. and Montagna, C. 2014. Chromosomal rearrangements in cancer: Detection and potential causal mechanisms. *Molecular & Cellular Oncology* 1(1), doi: 10.4161/mco.29904
- Hata, T. et al. 2018. Simple Detection of Telomere Fusions in Pancreatic Cancer, Intraductal Papillary Mucinous Neoplasm, and Pancreatic Cyst Fluid. *The Journal of Molecular Diagnostics* 20(1), pp. 46-55. doi: 10.1016/j.jmoldx.2017.09.006
- Hayes, M. 2019. Computational Analysis of Structural Variation in Cancer Genomes. *Methods in molecular biology (Clifton, N.J.)* 1878, p. 65. doi: 10.1007/978-1-4939-8868-6\_3
- Hayflick, L. 1965. The limited in vitro lifetime of human diploid cell strains. *Experimental cell research* 37(3), pp. 614-636.
- Hemann, M. T. et al. 2001. The Shortest Telomere, Not Average Telomere Length, Is Critical for Cell Viability and Chromosome Stability. *Cell* 107(1), pp. 67-77. doi: 10.1016/S0092-8674(01)00504-9
- Henson, J., D. et al. 2002. Alternative lengthening of telomeres in mammalian cells. *Oncogene* 21(4), p. 598. doi: 10.1038/sj.onc.1205058

- Her, J. and Bunting, S. F. 2018. How cells ensure correct repair of DNA double-strand breaks. *The Journal of biological chemistry* 293(27), p. 10502. doi: 10.1074/jbc.TM118.000371
- Herbig, U. et al. 2004. Telomere Shortening Triggers Senescence of Human Cells through a Pathway Involving ATM, p53, and p21 CIP1, but Not p16 INK4a. *Molecular Cell* 14(4), pp. 501-513. doi: 10.1016/S1097-2765(04)00256-4
- Higgins, G. S. et al. 2010. Overexpression of POLQ confers a poor prognosis in early breast cancer patients. *Oncotarget* 1(3), pp. 175-184. doi: 10.18632/oncotarget.124
- Hogg, M. et al. 2012. Promiscuous DNA synthesis by human DNA polymerase  $\theta$ . *Nucleic acids research* 40(6), p. 2611. doi: 10.1093/nar/gkr1102
- Hustedt, N. and Durocher, D. 2016. The control of DNA repair by the cell cycle. *Nature Cell Biology* 19(1), p. 1. doi: 10.1038/ncb3452
- Jafri, M. A. et al. 2016. Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Medicine* 8(1), doi: 10.1186/s13073-016-0324-x
- Jensen, R., B. et al. 2010. Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature* 467(7316), p. 678. doi: 10.1038/nature09399
- Jones, E. et al. 2001. SciPy: Open source scientific tools for Python.
- Jones, R. E. et al. 2014. Escape from Telomere-Driven Crisis Is DNA Ligase III Dependent. *Cell Reports* 8(4), pp. 1063-1076. doi: 10.1016/j.celrep.2014.07.007
- Kais, Z. et al. 2016. FANCD2 Maintains Fork Stability in BRCA1/2-Deficient Tumors and Promotes Alternative End-Joining DNA Repair. *Cell Reports* 15(11), pp. 2488-2499. doi: 10.1016/j.celrep.2016.05.031
- Kawamura, K. et al. 2004. DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. *International journal of cancer* 109(1), p. 9. doi: 10.1002/ijc.11666
- Kent, T. et al. 2015. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nature Structural & Molecular Biology* 22(3), pp. 230-237. doi: 10.1038/nsmb.2961
- Kent, T. et al. 2016. Polymerase theta is a robust terminal transferase that oscillates between three different mechanisms during end-joining. *Elife* 5, doi: 10.7554/eLife.13740

- Kielbasa, S. M. et al. 2011. Adaptive seeds tame genomic sequence comparison. *Genome research* 21(3), p. 487. doi: 10.1101/gr.113985.110
- Kim, N. W. et al. 1994. Specific association of human telomerase activity with immortal cells and cancer. *Science (New York, N.Y.)* 266(5193), p. 2011. doi: 10.1126/science.7605428
- Kim, W. et al. 2016. Regulation of the Human Telomerase Gene TERT by Telomere Position Effect-Over Long Distances (TPE-OLD): Implications for Aging and Cancer. *PLoS Biology* 14(12), p. e2000016. doi: 10.1371/journal.pbio.2000016
- Kimura, M. et al. 2010. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nature Protocols* 5(9), p. 1596. doi: 10.1038/nprot.2010.124
- Kotecki, M. et al. 1999. Isolation and Characterization of a Near-Haploid Human Cell Line. *Experimental Cell Research* 252(2), pp. 273-280. doi: <https://doi.org/10.1006/excr.1999.4656>
- Krokan, H. E. and Bjørås, M. 2013. Base excision repair. *Cold Spring Harbor perspectives in biology* 5(4), p. a012583. doi: 10.1101/cshperspect.a012583
- Lafferty-Whyte, K. et al. 2009. A gene expression signature classifying telomerase and ALT immortalization reveals an hTERT regulatory network and suggests a mesenchymal stem cell origin for ALT. *Oncogene* 28(43), p. 3765. doi: 10.1038/onc.2009.238
- Lalonde, M. and Chartrand, P. 2020. TERRA, a multifaceted regulator of telomerase activity at telomeres. *Journal of Molecular Biology*,
- Lamarche, B. J. et al. 2010. The MRN complex in double-strand break repair and telomere maintenance.
- Lansdorp, P. et al. 1996. Heterogeneity in telomere length of human chromosomes. *Human Molecular Genetics* 5(5), pp. 685-691.
- Lazzerini-Denchi, E. and Sfeir, A. 2016. Stop pulling my strings - what telomeres taught us about the DNA damage response. *Nature Reviews Molecular Cell Biology* 17(6), pp. 364-378. doi: 10.1038/nrm.2016.43
- Leao, R. et al. 2018. Mechanisms of human telomerase reverse transcriptase (hTERT) regulation: clinical impacts in cancer. *Journal Of Biomedical Science* 25(1), doi: 10.1186/s12929-018-0422-8
- Lei, M. et al. 2004. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature Structural & Molecular Biology* 11(12), p. 1223. doi: 10.1038/nsmb867

Lemée, F. et al. 2010. DNA polymerase  $\theta$  up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proceedings of the National Academy of Sciences* 107(30), p. 13390. doi: 10.1073/pnas.0910759107

Letsolo, B. T. et al. 2010. Fusion of short telomeres in human cells is characterized by extensive deletion and microhomology, and can result in complex rearrangements. *Nucleic Acids Research* 38(6), pp. 1841-1852. doi: 10.1093/nar/gkp1183

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14), pp. 1754-1760. doi: 10.1093/bioinformatics/btp324

Li, H. et al. 2009. The Sequence Alignment/ Map format and SAMtools. *Bioinformatics* 25(16), pp. 2078-2079. doi: 10.1093/bioinformatics/btp352

Liddiard, K. et al. 2018. DNA Ligase 1 is an essential mediator of sister chromatid telomere fusions in G2 cell cycle phase.

Liddiard, K. et al. 2016. Sister chromatid, but not NHEJ-mediated inter-chromosomal telomere fusions, occur independently of DNA ligases 3 and 4.

Lin, T. T. et al. 2010. Telomere dysfunction and fusion during the progression of chronic lymphocytic leukemia: evidence for a telomere crisis. *Blood* 116(11), pp. 1899-1907. doi: 10.1182/blood-2010-02-272104

Liu, D. et al. 2004. PTP interacts with POT1 and regulates its localization to telomeres. *Nature Cell Biology* 6(7), p. 673. doi: 10.1038/ncb1142

Liu, L. et al. 2012. Comparison of Next-Generation Sequencing Systems. *Journal Of Biomedicine And Biotechnology* 2012, doi: 10.1155/2012/251364

Liu, T. and Huang, J. 2016. DNA End Resection: Facts and Mechanisms. *Genomics, Proteomics & Bioinformatics* 14(3), pp. 126-130. doi: 10.1016/j.gpb.2016.05.002

Loayza, D. et al. 2004. DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites. *The Journal of biological chemistry* 279(13), p. 13241. doi: 10.1074/jbc.M312309200

Loeb, L., A. and Monnat, R., J. 2008. DNA polymerases and human disease. *Nature Reviews Genetics* 9(8), p. 594. doi: 10.1038/nrg2345

Luke, B. and Lingner, J. 2009. TERRA: telomeric repeat-containing RNA. *EMBO Journal* 28(17), pp. 2503-2510. doi: 10.1038/emboj.2009.166

- Lundblad, V. 2012. Telomere end processing: unexpected complexity at the end game. *Genes & development* 26(11), p. 1123. doi: 10.1101/gad.195339.112
- Lydeard, J., R. et al. 2007. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* 448(7155), p. 820. doi: 10.1038/nature06047
- Ma, Y. et al. 2005. The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair* 4(7), pp. 845-851. doi: 10.1016/j.dnarep.2005.04.013
- Maciejowski, J. and De Lange, T. 2017. Telomeres in cancer: tumour suppression and genome instability. *Nature Reviews Molecular Cell Biology* 18(3), doi: 10.1038/nrm.2016.171
- Malkova, A. and Ira, G. 2013. Break-induced replication: functions and molecular mechanism. *Current Opinion in Genetics & Development* 23(3), pp. 271-279. doi: 10.1016/j.gde.2013.05.007
- Mao, P. et al. 2016. Homologous recombination-dependent repair of telomeric DSBs in proliferating human cells. *Nature Communications* 7(1), doi: 10.1038/ncomms12154
- Mateos-Gomez, P. A. et al. 2015. Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature* 518(7538), pp. 254-U285. doi: 10.1038/nature14157
- Mateos-Gomez, P. A. et al. 2017. The helicase domain of Poltheta counteracts RPA to promote alt-NHEJ. *Nat Struct Mol Biol* 24(12), pp. 1116-1123. doi: 10.1038/nsmb.3494
- McClintock, B. 1941. The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics* 26(2), p. 234.
- Mcgee, J., S. et al. 2010. Reduced Rif2 and lack of Mec1 target short telomeres for elongation rather than double-strand break repair. *Nature Structural & Molecular Biology* 17(12), p. 1438. doi: 10.1038/nsmb.1947
- Mefford, H., C. and Trask, B., J. 2002. The complex structure and dynamic evolution of human subtelomeres. *Nature Reviews Genetics* 3(2), p. 91. doi: 10.1038/nrg727
- Mender, I. and Shay, J. W. 2015. Telomerase repeated amplification protocol (TRAP). *Bio-protocol* 5(22),
- Meyne, J. et al. 1990. Distribution of non-telomeric sites of the (TTAGGG)<sub>n</sub> telomeric sequence in vertebrate chromosomes. *Chromosoma* 99(1), pp. 3-10. doi: 10.1007/BF01737283

- Mi, H. et al. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* 47(D1), pp. D419-D426. doi: 10.1093/nar/gky1038
- Mitelman, F. et al. 1997. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Genetics* 15, pp. 417-474.
- Mladenov, E. et al. 2019. Strong suppression of gene conversion with increasing DNA double-strand break load delimited by 53BP1 and RAD52. *Nucleic acids research*, doi: 10.1093/nar/gkz1167
- Motycka, T. A. et al. 2004. Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *The Journal of biological chemistry* 279(14), p. 13634. doi: 10.1074/jbc.M313779200
- Moyzis, R. K. et al. 1988. A highly conserved repetitive DNA sequence, (TTAGGG)<sub>n</sub>, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* 85(18), p. 6622. doi: 10.1073/pnas.85.18.6622
- Muller, H. J. 1938. The remaking of chromosomes. Collecting net.
- Muraki, K. et al. 2015. Processing by MRE11 is involved in the sensitivity of subtelomeric regions to DNA double-strand breaks. *Nucleic acids research* 43(16), p. 7911. doi: 10.1093/nar/gkv714
- Muraki, K. and Murnane, J. P. 2017. The DNA damage response at dysfunctional telomeres, and at interstitial and subtelomeric DNA double-strand breaks. *Genes & Genetic Systems* 92(3), pp. 135-152. doi: 10.1266/ggs.17-00014
- Murnane, J. P. and Sabatier, L. 2004. Chromosome rearrangements resulting from telomere dysfunction and their role in cancer. Hoboken.
- Nassour, J. et al. 2019. Autophagic cell death restricts chromosomal instability during replicative crisis. *Nature* 565(7741), p. 659. doi: 10.1038/s41586-019-0885-0
- Newman, J. A. et al. 2015. Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway. *Structure* 23(12), pp. 2319-2330. doi: 10.1016/j.str.2015.10.014
- Nguyen, B. N. et al. 2009. Mechanism of dominant-negative telomerase function. *Cell Cycle* 8(19), pp. 3227-3233. doi: 10.4161/cc.8.19.9788
- Nick McElhinny, S. A. et al. 2000. Ku Recruits the XRCC4-Ligase IV Complex to DNA Ends. *Molecular and Cellular Biology* 20(9), p. 2996. doi: 10.1128/MCB.20.9.2996-3003.2000

- Nilsen, G. et al. 2012. Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. *13*(1), doi: 10.1186/1471-2164-13-591
- Nowell, P. C. and Hungerford, D. A. 1960. Chromosome studies on normal and leukemic human leukocytes. *Journal of the National Cancer Institute* 25(1), pp. 85-109. doi: 10.1093/jnci/25.1.85
- O'Sullivan, R., J. and Karlseder, J. 2010. Telomeres: protecting chromosomes against genome instability. *Nature Reviews Molecular Cell Biology* 11(3), p. 171. doi: 10.1038/nrm2848
- Ozdemir, A. Y. et al. 2018. Polymerase  $\theta$ -helicase efficiently unwinds DNA and RNA-DNA hybrids. *The Journal of biological chemistry* 293(14), p. 5259. doi: 10.1074/jbc.RA117.000565
- Panier, S. and Boulton, S. J. 2013. Double-strand break repair: 53BP1 comes into focus. *Nature Reviews Molecular Cell Biology* 15(1), p. 7. doi: 10.1038/nrm3719
- Pathania, S. et al. 2014. BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nature communications* 5(1), pp. 1-15.
- Pillaire, M. J. et al. 2009. A 'DNA replication' signature of progression and negative outcome in colorectal cancer. *Oncogene* 29(6), p. 876. doi: 10.1038/onc.2009.378
- Pokhrel, N. et al. 2017. Monitoring Replication Protein A (RPA) dynamics in homologous recombination through site-specific incorporation of non-canonical amino acids. *Nucleic acids research* 45(16), p. 9413. doi: 10.1093/nar/gkx598
- Quinlan, A. R. and Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6), pp. 841-842. doi: 10.1093/bioinformatics/btq033
- Richards, J. D. et al. 2008. Structure of the DNA repair helicase Hel308 reveals DNA binding and autoinhibitory domains. *Journal of Biological Chemistry* 283(8), pp. 5118-5126. doi: 10.1074/jbc.M707548200
- Riethman, H. et al. 2004. Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome research* 14(1), p. 18. doi: 10.1101/gr.1245004
- Robin, J. D. et al. 2014. Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. *Genes & development* 28(22), p. 2464. doi: 10.1101/gad.251041.114
- Robinson, J. T. et al. 2011. Integrative genomics viewer. *Nature Biotechnology* 29(1), p. 24. doi: 10.1038/nbt.1754

- Roger, L. et al. 2013. Extensive Telomere Erosion in the Initiation of Colorectal Adenomas and Its Association With Chromosomal Instability. *Jnci-Journal of the National Cancer Institute* 105(16), pp. 1202-1211. doi: 10.1093/jnci/djt191
- Rosenbloom, K. R. et al. 2015. The UCSC Genome Browser database: 2015 update. *Nucleic acids research* 43(Database issue), p. D670. doi: 10.1093/nar/gku1177
- Ryba, T. et al. 2012. Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome Research* 22(10), pp. 1833-1844. doi: 10.1101/gr.138511.112
- Ryba, T. et al. 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome research* 20(6), p. 761. doi: 10.1101/gr.099655.109
- Rybanska-Spaeder, I. et al. 2014. 53BP1 mediates the fusion of mammalian telomeres rendered dysfunctional by DNA-PKcs loss or inhibition. *PLoS ONE* 9(9), p. e108731. doi: 10.1371/journal.pone.0108731
- Sakharkar, M. K. et al. 2004. Distributions of exons and introns in the human genome. *In silico biology* 4(4), p. 387.
- Sambrook, J. et al. 1989. *Molecular cloning: a laboratory manual*. Cold spring harbor laboratory press.
- Schlacher, K. et al. 2011. Double-Strand Break Repair-Independent Role for BRCA2 in Blocking Stalled Replication Fork Degradation by MRE11. *Cell* 145(4), pp. 529-542. doi: 10.1016/j.cell.2011.03.041
- Schlacher, K. et al. 2012. A Distinct Replication Fork Protection Pathway Connects Fanconi Anemia Tumor Suppressors to RAD51-BRCA1/2. *Cancer Cell* 22(1), pp. 106-116. doi: 10.1016/j.ccr.2012.05.015
- Schmidt, J. and Cech, T. 2015. Human telomerase: biogenesis, trafficking, recruitment, and activation. *Genes & Development* 29(11), pp. 1095-1105. doi: 10.1101/gad.263863.115
- Scully, R. et al. 2019. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology* 20(11), pp. 698-714. doi: 10.1038/s41580-019-0152-0
- Sehgal, N. et al. 2014. Gene density and chromosome territory shape. *Chromosoma* 123(5), pp. 499-513. doi: 10.1007/s00412-014-0480-y

- Seki, M. et al. 2003. POLQ (Pol  $\theta$ ), a DNA polymerase and DNA-dependent ATPase in human cells. *Nucleic Acids Research* 31(21), pp. 6117-6126. doi: 10.1093/nar/gkg814
- Sfeir, A. et al. 2010. Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science (New York, N.Y.)* 327(5973), p. 1657. doi: 10.1126/science.1185100
- Sfeir, A. et al. 2009. Mammalian Telomeres Resemble Fragile Sites and Require TRF1 for Efficient Replication. *Cell* 138(1), pp. 90-103. doi: 10.1016/j.cell.2009.06.021
- Shamanna, R. A. et al. 2016. WRN regulates pathway choice between classical and alternative non-homologous end joining. *Nature communications* 7(1), p. 13785. doi: 10.1038/ncomms13785
- Shay, J., W. and Wright, W., E. 2000. Hayflick, his limit, and cellular ageing. *Nature Reviews Molecular Cell Biology* 1(1), p. 72. doi: 10.1038/35036093
- Shima, N. et al. 2003. Phenotype-based identification of mouse chromosome instability mutants. *Genetics* 163(3), pp. 1031-1040.
- Shima, N. et al. 2004. The Mouse Genomic Instability Mutation chaos1 Is an Allele of Polq That Exhibits Genetic Interaction with Atm. *Molecular and Cellular Biology* 24(23), p. 10381. doi: 10.1128/MCB.24.23.10381-10389.2004
- Shlien, A. and Malkin, D. 2009. Copy number variations and cancer. *Genome Medicine* 1(6), doi: 10.1186/gm62
- Smith, J. et al. 2000. Characterization of genetic interactions with RFA1: the role of RPA in DNA replication and telomere maintenance. *Biochimie* 82(1), pp. 71-78. doi: 10.1016/S0300-9084(00)00183-8
- Stewénus, Y. et al. 2007. Defective chromosome segregation and telomere dysfunction in aggressive Wilms' tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research* 13(22 Pt 1), p. 6593. doi: 10.1158/1078-0432.CCR-07-1081
- Surrallés, J. et al. 1997. Chromosomes with high gene density are preferentially repaired in human cells. 12(6), pp. 437-442. doi: 10.1093/mutage/12.6.437
- Symington, L. S. and Gautier, J. 2011. Double-Strand Break End Resection and Repair Pathway Choice. *Annual Review of Genetics* 45(1), pp. 247-271. doi: 10.1146/annurev-genet-110410-132435

- Tabach, Y. et al. 2011. Amplification of the 20q Chromosomal Arm Occurs Early in Tumorigenic Transformation and May Initiate Cancer. *PLoS ONE* 6(1), doi: 10.1371/journal.pone.0014632
- Takai, Kaori k. et al. 2011. Telomere Protection by TPP1/POT1 Requires Tethering to TIN2. *Molecular Cell* 44(4), pp. 647-659. doi: 10.1016/j.molcel.2011.08.043
- Tanaka, H. et al. 2012. Telomere fusions in early human breast carcinoma. *Proceedings of the National Academy of Sciences* 109(35), p. 14098. doi: 10.1073/pnas.1120062109
- Tanaka, H. et al. 2014. The Presence of Telomere Fusion in Sporadic Colon Cancer Independently of Disease Stage, TP53/KRAS Mutation Status, Mean Telomere Length, and Telomerase Activity. *Neoplasia* 16(10), pp. 814-823. doi: 10.1016/j.neo.2014.08.009
- Tashiro, S. et al. 2017. NAR breakthrough article: Subtelomeres constitute a safeguard for gene expression and chromosome homeostasis. *Nucleic Acids Research* 45(18), pp. 10333-10349. doi: 10.1093/nar/gkx780
- Thanasoula, M. et al. 2012. ATM/ATR checkpoint activation downregulates CDC25C to prevent mitotic entry with uncapped telomeres. *EMBO Journal* 31(16), pp. 3398-3410. doi: 10.1038/emboj.2012.191
- Truong, L., N. et al. 2013. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proceedings of the National Academy of Sciences* 110(19), p. 7720. doi: 10.1073/pnas.1213431110
- Umbreit, N. et al. 2019. Mechanisms Generating Cancer Genome Complexity From A Single Cell Division Error. *bioRxiv*, p. 835058.
- van Steensel, B. et al. 1998. TRF2 Protects Human Telomeres from End-to-End Fusions. *Cell* 92(3), pp. 401-413. doi: 10.1016/S0092-8674(00)80932-0
- Victorelli, S. and Passos, J. 2017. Telomeres and Cell Senescence - Size Matters Not. *Ebiomedicine* 21(C), pp. 14-20. doi: 10.1016/j.ebiom.2017.03.027
- Von Zglinicki, T. 2002. Oxidative stress shortens telomeres. *Trends in Biochemical Sciences* 27(7), pp. 339-344. doi: 10.1016/S0968-0004(02)02110-2
- Wang, H. et al. 2003. Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic acids research* 31(18), p. 5377. doi: 10.1093/nar/gkg728
- Wang, M. et al. 2006. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic acids research* 34(21), p. 6170. doi: 10.1093/nar/gkl840

- Wenz, C. et al. 2001. Human telomerase contains two cooperating telomerase RNA molecules. *EMBO Journal* 20(13), pp. 3526-3534. doi: 10.1093/emboj/20.13.3526
- Wold, M. S. 1997. REPLICATION PROTEIN A:A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. *Annual Review of Biochemistry* 66(1), pp. 61-92. doi: 10.1146/annurev.biochem.66.1.61
- Wright, W. E. and Shay, J. W. 1992. The two-stage mechanism controlling cellular senescence and immortalization. *Experimental Gerontology* 27(4), pp. 383-389. doi: 10.1016/0531-5565(92)90069-C
- Wyatt, D. W. et al. 2016. Essential Roles for Polymerase theta-Mediated End Joining in the Repair of Chromosome Breaks. *Molecular Cell* 63(4), pp. 662-673. doi: 10.1016/j.molcel.2016.06.020
- Xin, H. et al. 2008. The telosome/shelterin complex and its functions. *Genome Biology* 9(9), doi: 10.1186/gb-2008-9-9-232
- Yousefzadeh, M. J. and Wood, R. D. 2013. DNA polymerase POLQ and cellular defense against DNA damage. *DNA Repair* 12(1), pp. 1-9. doi: 10.1016/j.dnarep.2012.10.004
- Yousefzadeh, M. J. et al. 2014. Mechanism of Suppression of Chromosomal Instability by DNA Polymerase POLQ. *Plos Genetics* 10(10), doi: 10.1371/journal.pgen.1004654
- Yuan, X. et al. 2019. Mechanisms underlying the activation of TERT transcription and telomerase activity in human cancer: old actors and new players. *Oncogene* 38(34), pp. 6172-6183. doi: 10.1038/s41388-019-0872-9
- Zahn, K. E. et al. 2011. Phosphonoformic acid inhibits viral replication by trapping the closed form of the DNA polymerase. *Journal of Biological Chemistry* 286(28), pp. 25246-25255. doi: 10.1074/jbc.M111.248864
- Zhang, J.-M. et al. 2019. Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. *Cell Reports* 26(4), pp. 955-968.e953. doi: 10.1016/j.celrep.2018.12.102
- Zhao, W. et al. 2017. BRCA1–BARD1 promotes RAD51-mediated homologous DNA pairing. *Nature* 550(7676), p. 360. doi: 10.1038/nature24060