Department of Cancer and Genetics
School of Medicine
Cardiff University

An analysis of telomere dynamics in breast cancer

A thesis submitted to the School of Medicine, Cardiff University in partial fulfilment for the degree of Doctor of Philosophy

By
Kate Simpson
2013

V1

Publications:

Work submitted in this body of work has been submitted in a paper, and is at present under review for publication.

Acknowledgments

Firstly I would like to thank my brilliant supervisor Professor Duncan Baird, without whose continuing support and guidance I couldn't have completed this body of work. I am also very grateful to have had the support and friendship of everyone in the Cancer and Genetics lab: Julia grimstead, Rhiannon Robinson, Sophie Betteridge, Megan Musson, Kevin Norris, Maira Tankimanova, Thet Thet Lin, Boitelo Letsolo, Laureline Roger, Nicole Heppel, Bethan Britt-Compton, Jenna Williams, Kate Liddiard and Jenna Williams.

I would like to thank my partner Caradog Rogers for being my best friend throughout these three years, and for keeping me sane! I would also like to thank my family for their interest and enthusiasm in my work: Glyn Pope and Jill Pope and Chris Simpson.

Summary

Telomeres are specialised structures that that cap the ends of chromosomes; and prevent the natural end of a chromosome from being recognised as a double stranded DNA break. Telomeres erode with ongoing cell division ultimately leading the induction of replicative senescence that provides a proliferative lifespan barrier. In the absence of a functional DNA damage response, telomeres are prone to end-fusions, creating dicentric chromosomes that can initiate breakage fusion–bridge cycles. Evidence from studies in well-defined mouse models of cancer, and also human tumours, have shown that telomere dysfunction may be a key event in the progression of disease via the increasing genomic instabilities that arise from telomere fusion events.

The key aim of this thesis was to test the hypothesis that telomere dysfunction occurs during the progression of breast cancer and that this can drive the large-scale genomic rearrangement frequently observed in this disease.

Technology development was attempted in the hope to allow single-molecule telomere fusion detection of more complex mutational structures, with limited success but possible future potential.

Telomere length analysis was carried out using high resolution single molecule PCR strategies (STELA) in a panel of DNA samples derived from invasive ductal carcinoma. Telomere length data analysed alongside clinical data received for all samples was used for the potential utility of telomere length as a potential prognostic indicator in breast cancer.

A key finding of this work was to demonstrate the use STELA combined with statistical tests to show that short telomere length is highly significant in terms of prognosis in breast cancer. Telomere length stratification could thus potentially be used as a method of defining new breast cancer subtypes in terms of severity.

Contents Page

Chapter 6: an Analysis of Telomere fusion in Invasive Ductal Carcinoma

Chapter 7: General Discussion

Abbreviations

A Adenosine

ADH Atypical ducatl hyperplasia

AKT v-akt murine thymoma viral oncogene homolog 1

ALL Acute lymphocytic leukaemia

ALM Adaptor-Ligation Mediated

ALT Alternative lengthening of telomeres

AML Acute myeloid leukaemia

AMT Aminomethyltransferase

APB ALT-associated PML bodies

ATM Ataxia telangiectasia mutated protein

ATR Ataxia telangiectasia and Rad3 related

ATRX Alpha thalassemia/mental retardation syndrome X-linked

A-CGH Array- Comparative genomic hybridisation

BC Breast cancer

BFB Breakage-Fusion-Bridge

BRCA1 Breast Cancer Susceptibility 1

BRCA2 Breast Cancer Susceptibility 2

BP Base pair

C Cytosine

CA Chromosomal aberration

CGH Comparative genomic hybridisation

CHK1 Checkpoint kinase 1

CHK2 Checkpoint kinase 2

CI Chromosomal Instability

CLL Chronic lymphocytic leukaemia

CO-FISH Chromosome orientation-FISH

ChIP Chromatin immuneprecipitation

cDNA complementary DNA

CNA Copy number abberation

CNV Copy number variation

DCIS Ductal carcinoma *in situ*

DDR DNA damage response/repair

DMSO Dimethylsulfoxide

DNA Deoxyribonucleic acid

DNA-PK DNA-protein kinase

DNA-PKcs DNA-protein kinase catalytic-subunit

DNA-PKcsi DNA-protein kinase catalytic-subunit inhibitor

DSB Double strand breaks

dsDNA double stranded DNA

DW Distilled water

ER Oestrogen receptor

ERBB2 Erythroblastosis oncogene B

FISH Fluorescence *in-situ* hybridization

gDNA genomic DNA

G Guanine

HER2 Human Epidermal Growth Factor Receptor 2

HNPCC Hereditary non-polyposis colorectal carcinoma

HR Homologous recombination, Hazard ratio or Hormone receptor depending on the context used

H2AX H2A histone family, member X

IDC Invasive ductal carcinoma

IF Immunofluorescence

IF-TIF Immunofluorescence-Telomere dysfunction induced foci

ILC Invasive lobular carcinoma

IR Ionising Radiation

KB Kilobase

LOF Loss of function

LOH loss of heterozygosity

LR Local recurrence

MAPK Mitogen-activated protein kinase

MEF Mouse embryonic fibroblasts

miRNA microRNA

MRN complex consisting of Mre11-Rad-50-NSB1

MSH2 MutS homolog 2, colon cancer, nonpolyposis type 1

MSH3 MutS homolog 2, colon cancer, nonpolyposis type 3

MSH6 MutS homolog 2, colon cancer, nonpolyposis type 6

MYC V-myc myelocytomatosis viral oncogene homolog (avian)

NHEJ Non-homologous end joining

NRT Non reciprocal translocation

NSB1 Nijmegen breakage syndrome 1

PARP Poly (ADP-ribose) polymerase

PCR Polymerase chain reaction

PD population doubling

PGR/PR Progesterone receptor

PI3K phosphoinositide 3-kinase

PML Promyelocytic leukemia protein

PNA Peptide nucleic acid

POT1 Protection of telomeres 1

PR Progesterone receptor

PTEN Phosphatase and tensin homolog

Q-FISH Quantitative in-situ hybridization

RAD 50 DNA repair protein RAD50

RAD 51 DNA repair protein RAD51

RAP1 Repressor activator protein 1

RNA Ribonucleic acid

RNAi Ribonucleic acid interference

RT-PCR Reverse transcriptase-polymerase chain reaction

SCE Sister chromatid exchange

SD Standard deviation

SDS Sodium dodecyl sulphate

SEM Standard error of mean

siRNA Short interfering ribonucleic acid

SSC Sodium chloride sodium citric acid

ssDNA single stranded DNA

STELA Single Telomere Length Analysis

T Thymidine

T-loop Telomeric-loop

Telo-FISH Telomeric fluorescence in-situ hybridization

TERT Telomerase reverse transcriptase

TERC Telomerase RNA component

TFI Telomeric fluorescence intensity

TIF Telomere dysfunction induced foci

TIN2 TRF1-interacting factor

TLV Telomere Length Variation

TNM Tumour node metastasis

TPE Telomere postion effect

TPP1 Tripeptidyl peptidase I

TP53 Tumour protein 53

TRF Terminal Restriction Fragment

TRF1 Telomeric repeat binding factor 1

TRF2 Telomeric repeat binding factor 2

T-SCE Telomere-sister chromatid exchange

TVR Telomere variant repeat

UDH Usual ductal hyperplasia

UV Ultra violet

WRN Werner syndrome ATP-dependent helicase

Chapter 1: Introduction

1.1: History of Telomere Biology

The name ‘Telomere’ stems from the Greek ‘Telos’ meaning end, and ‘Meros’ meaning part. This term was first used in 1938 by North American geneticist Hermann J. Müller (Müller, 1938). He observed that during the irradiation of *Drosophila melanogastor* genomes with X-rays, the DNA was subjected to a plethora of gene mutations and chromosomal rearrangements resulting from DNA breaks, such as translocations, inversions and deletions (Müller, 1938). The termini of these irradiated chromosomes however, did not exhibit any such alterations. He thus speculated of the presence of a ‘terminal gene’ that protects the ends of DNA (Müller, 1938; Blackburn and Greider, 1996)

Barbara McClintock, in 1939, was working on *Zea Mays* cytogenetics and discovered that when the terminal structures of DNA are lost during X-irradiation, sister chromatids tend to fuse with one another forming dicentric chromsomes (McClintock, 1941). When the sister chromatids are pulled apart during meiosis an anaphase bridge is formed that can break at spontaneous sites (McClintock, 1941). Cycles of this ‘breakage-fusion-bridge’ (BFB) process (Figure 1.1) can result in unequal exchange of genetic information and several rounds of meiosis in this way can lead to chromosomal aberrations such as gene amplification (McClintock, 1941). Barbara McClintock demonstrated that in somatic cells these cycles of BFB can continue indefinitely, but in germ cells, regardless of this damage, the ends could be restored thanks to the acquisition of new telomere (McClintock, 1941).

In 1965 Leonard Hayflick published research into ageing showing that human diploid cells have a limited replicative lifespan and cannot divide indefinitely Hayflick and Moorhead,

1961). These cells appeared to have a cell-intrinsic mechanism that was capable of counting the number of cell divisions that they had undergone (Hayflick and Moorhead, 1961).

In the 1970's James Watson and Alexey Olovnikov worked independently on the 'end replication problem'; how in linear eukaryote chromosomescannot be fully replicated to the terminus of the DNA sequence which would result in a loss of DNA (Watson, 1972; Olovnikov, 1971). To solve this and to accommodate Hayflick's idea about limited somatic cell division (Hayflick and Moorhead, 1961), Olovnikov suggested that DNA sequences would be lost in every replicative phase until they reached a critical level, at which point cell division would stop and thus the loss of terminal sequences may provide a cell division counting mechanism (Olovnikov, 1973).

It wasn't until 1978, that Elizabeth Blackburn and her colleagues, sequencing *Tetrahymena thermophilia* chromosomes, discovered the presence of telomere repeat DNA (Blackburn et al., 1981).

Figure 1.1: Mechanism of BFB Cycles as described by Barbara McClintock

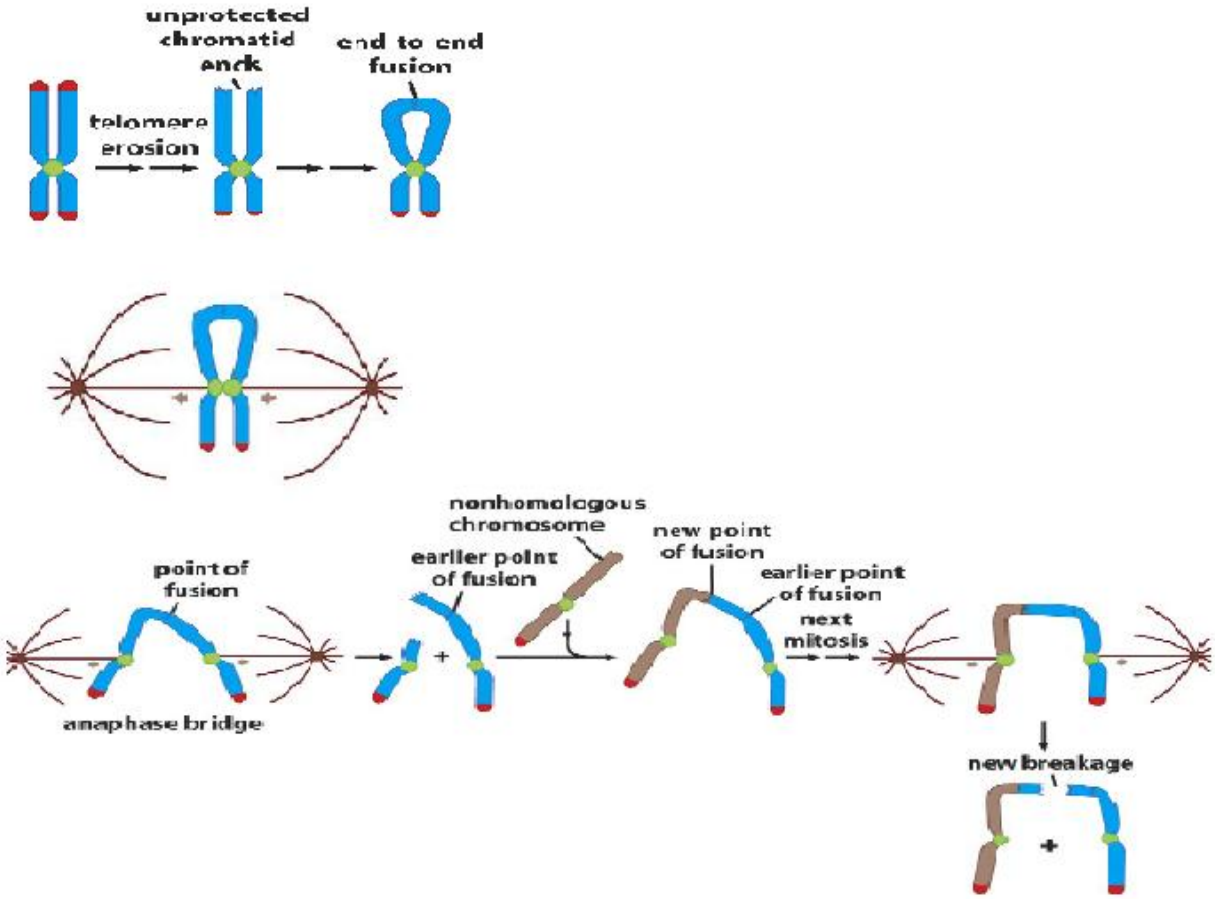


Figure 1.1: The way in which telomere erosion and subsequent fusion can cause cycles of anaphase bridging-fusion-breakage (BFB) and thus an array of chromosomal rearrangements at mitosis. Figure adapted from Gisselsson D . Chromosomal Instability in Cancer: Causes and Consequences. Atlas Genet Cytogenet Oncol Haematol. May 2001 .
URL : http://AtlasGeneticsOncology.org/Deep/ChromosomInstabilID20023.html

1.2: Telomere Structure

Telomeres consist of tandem repetitive DNA sequences located at the termini of linear chromosomes of most eukaryotic organisms. Telomeric DNA is highly conserved amongst eukaryotic species regardless of differences in length and sequence (Moyzis, 1988). Telomeres tend to consist of a G-rich strand at the 3′ end of the chromosome that overhangs a C-rich strand at the 5′ end. This G-strand 'overhang' was first observed in the hypotrichous ciliate organism *Oxtricha* whereby it was shown to be longer than the C-strand at telomere termini in this organism (Klotbutcher *et al.,* 1981). The G-strand overhang varies in size between species, with a length of 16 nt in *Oxytricha,* to a length that varies between 50–100 nt in mouse and human telomeres (Kipling and Cooke, 1990; DeLange et al., 1990). Mutations affecting the G-strand overhang in yeast (*Saccharomyces cerevisiae*) have shown to disrupt telomere function and thus it is considered to be essential for telomere function (**Polotnianka et al. 1998**). G-strand overhangs are maintained throughout the replicative lifespan of human cells (Chai *et al.,* 2005) either by elongation by telomerase or degradation of the C-strand. The G-strand overhang has been hypothesised to be present as a result of the end-replication problem. During mammalian replication, lagging strand synthesis involves RNA priming events that are considered to occur around every 100–600bp (Anderson and DePamphilis 1979). Links have been suggested between this end replication phenomenon and with the rates of telomere shortening of 40–200bp per cell in human cells *in vitro*.

1.2.1 Prokaryotes

The majority of prokaryotes chromosomes are arranged as circular structures (Charlebois,

1999) and this is how they could historically be easily distinguished from eukaryotic organisms (Mason & Powelson, 1956; Cairns, 1963, Charlebois, 1999) . This means that the end replication problem does not apply here. However, further studies have revealed has revealed that not all prokaryotes follow this pattern, some have multiple circular chromosomes of different sizes, such as *Rhodobacter spheroides* (Suwanto and Kaplan, 1989) and other have linear chromosomes, such as *Borrelia* (Casjens, 1999) or a mix of linear and circular such as a cluster of species in the genus *Streptomyces* (Chen, 1996). These prokaryotic organisms contain linear chromosome structures that also possess telomere-like structures. Some species of bacteria, such as some Proteobacteria, have palindromic hairpin loops that act to protect the DNA terminus (Jumas-Bilak et al. 1998); others have 'invertron' telomeres are protected by proteins that bind to the 5′-ends (Suwanto, A., and S. Kaplan. 1989).

1.2.2 Eukaryotes

Telomere sequence and length varies greatly between eukaryotic species, from approximately 300 to 600bp in yeast to many kilobases (5-20kb) in humans, and usually is composed of arrays of G-rich, six- to eight-base-pair-long repeats (Blackburn et al., 1984).

1.2.3 Diptera

Research into telomere structure in the species *Drosophila melanogastor* have shown Diptera to have a unique telomere structure compared with other orders, lacking telomerase and thus lacking the short arrays of simple repeats observed in most other organisms. Instead of these short repeat structures, *Drosophila* species have arrays of long repeats (6-10kb) or

telomere-specific long tandem arrays of two non-LTR (long terminal repeat) retrotransposons (Pardue and Debaryshe, 2003).

1.2.4 Protozoa

The *Tetrehymena thermophila* telomere structure was first characterised by Elizabeth Blackburn's laboratory (Blackburn and Gall, 1978) and has repeat arrays of 20-70 bps of TGGGG. This ciliate protozoan is a model organism in experimental biology due to it being easily cultured in abundance. *Trypanosomes* have been shown to have similar telomere sequences to numerous higher eukaryotes. Protazoa such as *Trypanosoma brucei* have an abundance of telomeres that can be easily studied due to the presence of a number of minichromosomes. *T. brucei* and *T.cruzi* telomeres consist of 10–20 kb of hexameric TTAGGG repeat arrays and a 3′ TTAGGG overhang (Blackburn and Challoner, 1984; **Stansel** *et al.,* 2001). Electron microscopy of *Trypanosomes* has revealed the presence of t-loops at the Telomeric termini, although these t-loops have been shown to be smaller than mammalian t-loops (~1kb instead of 25kb in mammals), suggesting that t-loops are conserved among eukaryotic organisms (**Stansel** *et al.,* 2001).

1.2.5 Fungi

Yeast telomeres tend to share similarity in structure and functionality to a lot of eukaryotic organisms in the way that both budding and fission yeast families have a G-rich terminus and are maintained by the action of telomerase to protect the end from degradation. The telomeric DNA sequences found in the budding yeast *Saccharomyces cerevisiae* are $TG_{2-3}(TG)_{1-6}$ repeat arrays and span around 120-150bp (**Cohn and Blackburn**., 1995; de Bruin *et al.*, 2001). In *S. cerevisiae* two arrays of telomere repeat DNA form a higher order structure different from t-loops found in other eukaryotes. Not all *Saccharomyces* telomeres share this pattern of

telomere array, *S. kluyveri* telomere DNA sequences are homogeneous 26-base pair repeats of GGGTGGACATGCGTACTGTGAGGTCT, and S. Castellii telomeric DNA consists of TCTGGGTG repeats (76%) interspersed with TCTGGG$(TG)_{2-4}$ repeats (24%) (**Cohn and Blackburn., 1995** and **Cohn et al. 1998**). *Schizosaccharomyces pombe*, a fission yeast, has the telomere array $G_{2–8}$TTAC(A) and is about 300bp long (Murray **and Szostak,** 1986).

1.2.6 Plants

Telomeric DNA from *Arabidopsis* was cloned in 1988 revealing repeat arrays of TTTAGGG (Richards and Ausubel, 1988), common to most angiosperms. No telomere-like sequence has been characterised in *Allium cepa* (onion) and other *Allium* species (Pich *et al.*, 1996). Other *Asparagales* species tend to have the hexameric repeat array TTAGGG, the same sequence as mammalian telomeres (Adams et al., 2001). Some *Asparagales* species have been observed to have variant tandem repeats such as TTGGGG and other species have a seven base tandem repeat of TTTAGGG (Sykorova et al., 2003). T-loops have been observed in plants, for example in *Pitum sativum* (pea plant) the average size of the t-loop is 22kb, but can measure up to 50kb (Cesare et al., 2003).

Table 1.1: Telomere repeats found in different organisms

Group	Organism	Telomeric repeat (5′ to 3′)
Vertebrates	Human, mouse, *Xenopus*	TTAGGG
Filamentous fungi	*Neurospora crassa*	TTAGGG
Slime moulds	*Physarum, Didymium*	TTAGGG
	Dictyostelium	AG(1-8)
Kinetoplastid protozoa	*Trypanosoma, Crithidia*	TTAGGG
Ciliate protozoa	*Tetrahymena, Glaucoma*	TTGGGG
	Paramecium	TTGGG(T/G)
	Oxytricha, Stylonychia, Euplotes	TTTTGGGG
Apicomplexan protozoa	*Plasmodium*	TTAGGG(T/C)
Higher plants	*Arabidopsis thaliana*	TTTAGGG
Green algae	*Chlamydomonas*	TTTTAGGG
Insects	*Bombyx mori*	TTAGG
Roundworms	*Ascaris lumbricoides*	TTAGGC
Fission yeasts	*Schizosaccharomyces pombe*	TTAC(A)(C)G(1-8)
Budding yeasts	*Saccharomyces cerevisiae*	TGTGGGTGTGGTG (from RNA template) or G(2-3)(TG)(1-6)T (consensus)
	Saccharomyces castellii	TCTGGGTG
	Candida glabrata	GGGGTCTGGGTGCTG
	Candida albicans	GGTGTACGGATGTCTAACTTCTT
	Candida tropicalis	GGTGTA[C/A]GGATGTCACGATCATT
	Candida maltosa	GGTGTACGGATGCAGACTCGCTT
	Candida guillermondii	GGTGTAC
	Candida pseudotropicalis	GGTGTACGGATTTGATTAGTTATGT
	Kluyveromyces lactis	GGTGTACGGATTTGATTAGGTATGT

Table 1.1 : The types of telomere repeat array that can be found are shown by group and subsequent examples of organisms for that group.

1.3: Mammalian Telomere structure

Mammalian telomeres consist of the highly conserved hexameric (TTAGGG)n tandem repeat DNA sequence (Meyne *et al.,* 1989). The length of telomere repeat arrays is species specific and genetically determined. In mice telomere repeat arrays vary between sub species and can extend up to 65 kb in length in *Mus musculus* and can be much shorted in length in wild *Mus spretus* being anywhere between 5-20kb which is the same as human telomeric length (Kipling and Cooke 1990; Hemann and Greider NAR 2000).

Human telomeres consist of lengths that average between 5-20kb of TTAGGG repeats and also degenerate sequence variants such as TCAGGG, TTGGGG and TGAGGG in the proximal 1-2kb (Lange *et al.,* 1990; Baird *et al.,* 2005).

1.4 Human Subtelomeric regions

Often telomeres and chromosome specific sequences are separated by a repetitive region known as a subtelomeric region. The subtelomere tends to contain a large number of segmental duplications, and due to the nature of them sharing high sequence identity it has been postulated that these subtelomeres evolved via recombinational mechanisms between chromosomes such as homologous recombination and non-homologous end joining (Richardson *et al.,* 1998). Rudd and colleagues found an enrichment of sister chromatid exchanges (SCE) in subtelomeres implicating subtelomeres as hotspots for DNA repair and exchange (Rudd., 2007). The segmental duplications can be unique to subtelomeres or shared between interstitial sites such as 2q13-q14. Gene families have been identified that share subtelomeric sequence homology, such as the olfactory receptor genes and zinc finger genes (Trask *et al.,* 1998 and Riethman *et al.,* 2001). A number of the duplications found at the subtelomere are situated in proximity to the terminus of multiple chromosome ends giving rise to subtelomere families sharing homology. Hybridization experiments using the

TelBam11 probe revealed a family of telomere sequences sharing a large degree of subtelomeric homology 21q, 1q, 2q, 5q, 6q, 6p, 8p, 10q, 13q, 17q, 19p, 19q, 22q and the 2q13 interstitial telomeric locus. Similarly using the TelBam3.4 probe revealed subtelomeres sharing homology for 16p, 1p, 9p, 12p, 15q, XqYq ant the 2q14 interstitial locus (Riethman *et al.,* 2001). These groups of telomeres have since been termed the 21q and 16p families respectively (Letsolo *et al.,* 2010).

1.4.1: Humans: The T-loop and G-strand overhang

Human telomeres consist of TTAGGG tandem repeats and end with a 3′ single stranded overhang (figure 1.1), which loops back forming a 'T-loop' lariat structure (Griffith et al., 1999). This 3′ overhang invades the double stranded DNA forming a region of single stranded DNA that is displaced (a displacement or D-loop) (Griffith et al., 1999; **Greider**, 1999).

The T-loop is held together by six core proteins: TRF1, TRF2, TIN2, POT1, TPP1 and RAP1 collectively referred to as the shelterin complex (DeLange, 2005). TRF1and TRF2 bind directly to double stranded DNA (van Steensel and de Lange, 1997); POT1 binds to the G-rich single stranded overhang, Rap1 and TPP1 bind to telomere indirectly via TRF1 and 2 (DeLange, 2005).

Figure 1.2: The shelterin complex and its interaction with the T-loop

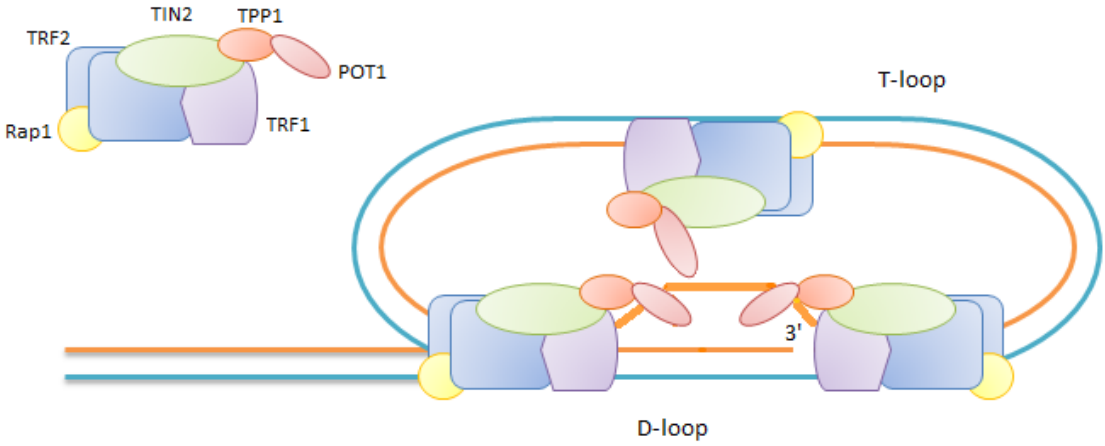


Figure 1.2 : Structure of the shelterin complex and its associated protein components. Adapted from 'Telomere biology in healthy ageing and disease' by Hisko et al. 2010; European Journal of Physiology Vol. 459 Issue 2.

1.3.2: The Shelterin Complex and associated proteins

1.4.2 TRF1 and TRF2

Telomeric structure is maintained by the interaction of several telomere-associated proteins. Telomere repeat-binding factor (TRF-1) is a protein that in humans is encoded by the *TERF1* gene (Shen et al., 1997). TRF1 was discovered first in HeLa cells when isolating a sequence that shared specificity with telomeric repeats (Zhong et al. 1992; Chong et al. 1995). TRF2 was isolated a few years later and is a protein that is related to TRF1 and is homologous in its structure and parts of its sequence (Bilaud et al. 1997; Broccoli et al. 1997). Both TRF1 and TRF2 bind double stranded TTAGGG repeats at the telomere as homodimers (DeLange, 2005; Court et al. 2005). This homodimerisation is regulated by the TRF-homology domain (TRFH) (Court et al. 2005). Regardless of these related features, TRF1 and TRF2 do not interact with one another, and their specific functions at the telomere appear to be quite different (DeLange, 2005).

TRF1 has been shown to exhibit DNA remodelling activity, (Bianchi et al. 1997; Griffith et al. 1998). This may contribute to the folding of the T-loop structure due to evidence that TRF1 can bend and pair telomere repeat arrays (Bianchi et al. 1997, 1999; Griffith et al.1998). TRF1 has been postulated as a negative regulator of telomere length by directly inhibiting telomerase (Kim et al. 2004; Xin, H et al. 2008). Conversely TRF2 has an important role in telomere cap protection (Kim et al. 2004). TRF2 inhibition results in the loss of single stranded telomeric repeats (TTAGGG) and widespread telomere fusion (Zhu et al. 2003).

1.4.3 The TRF1 Complex: Tankyrase 1 and 2, TIN2, PINX1 and POT1

As mentioned earlier Telomerase Repeat Binding Factor 1 (TRF1) is involved in telomeric structure and binds double stranded DNA at the telomere. Inhibition of TRF1 occurs by the interaction of poly ADP-ribose polymerases (PARP) proteins Tankyrase 1 and 2 (**Smith** et al., 1998; Cook et al., 2002). TRF1 has been shown to negatively regulate telomere length and

has been thought to do this by controlling the action of telomerase (Van Steensel et al., 2000). Tankyrase 1 and 2 ribosylate and thus inhibit the binding ability of TRF1 (Cook et al., 2002). Overexpression of tankyrase 1 has the same resulting phenotype as with the inhibition of TRF1 i.e. the elongation of telomeres. TIN2 (or TRF1 and TRF2-interacting protein 2) has a TRF1 binding domain at its C-terminal and has been shown to interact with both TRF1 and the tankyrase proteins, having a stabilisation effect on the complex of interactions (Cook et al., 2002). PINX1 has been shown to exhibit inhibitory effects on telomerase *in vitro* (Zhou and Lu, 2001; Zhou et al., 2011) and has thus been hypothesised to have a function in telomere length maintenance (Zhou et al., 2011).

POT1 (or Protection of Telomeres 1) is recruited to telomeres and bind to the single stranded 3′ overhang blocking telomerase from binding to telomeric ends, and thus acts as a telomeric length regulator (Baumann et al., 2002).

1.4.4 Rap1

The human form of Rap1 (Repressor/activator protein 1) was initially detected in connection to TRF2 and shares homology with *S.cerevisiae* scRap1 (Li et al., 2000). ScRap has been shown in yeast to bind directly to telomeric termini and control length regulation (Conrad et al., 1990; Lustig et al., 1990; Krauskopf and Blackburn, 1996). The human form (hRap1) binds to telomeric DNA via TRF2 interactions. Rap1 has a BRCT domain, a myb motif, a coiled region and a C-terminus (called an RCT domain) which mediates binding with TRF2 and thus the telomere itself (Li et al., 2000). Deletion of Rap1 in mice has shown a potential role in telomere length homeostasis and recombination (Lange *et al.,* 2010; Blasco *et al.,* 2010). Recent findings reveal that in mammals with deleted Rap1, there is a higher frequency of short telomeres and sister telomere recombination but did not show an increase in telomere fusions (Martinez et al., 2010). Further work also suggested that Rap1 was vital for the repression of homology-directed repair in mice, which can cause alterations in telomere

length (Lange *et al.,* 2010). This was shown by an increase in telomere sister chromatid exchanges in the absence of Rap1 (deLange *et al.,* 2010).

1.4.5 TPP1

Research in both mammals and the fission yeast *S. pombe* has postulated that the TPP1 protein binds to POT1 and is actually required for the POT1 mediated binding and subsequent protection of telomeres (Miyoshi *et al.*,2008).

Mouse models with deficient TPP1 show cell-cycle arrest, an increase in telomere shortening and damage (by the detection of telomere damage foci or TIFs) (Tejera et al., 2010) and decrease in the binding of telomerase subunit TERT (DeLange et al., 2009). Findings thus point towards TPP1 as having a role in both the protection of telomeres and the elongation of telomeres (DeLange et al., 2009).

1.4.6 Other telomere associated proteins

Many of transient telomere associated proteins are involved in DNA damage repair.

1.4.6.1 PARP 1 and PARP2

PARP1 and PARP2 are poly(ADP-ribose) polymerases that have similar functions in the repair of dsDNA (Kraus and Lis 2003). Cells deficient in both PARP1 and PARP2 have shown an increase in the occurrence of spontaneous genomic instability *in vitro* (Tong et al., 2001). PARP1 has also been implicated in DSB repair via homologous recombination (HR) (Waldman and Waldman, 1991), inhibition of PARP1 has been shown to suppress BRCA1 and RAD51 leading to inhibition of HR (Bryant et al. ,2005).

1.4.7 Rad 51D

Rad 51 is a protein family that are involved in the repair of double stranded (ds) DNA breaks (Pittman et al., 1998). One member of this protein family is Rad51D that has been shown to be involved in the repair of dsDNA breaks also have a role in telomere protection via

stabilising the terminal lariat structure T-loop (Tarsounas *et al.,* 2004). In mouse models with deletion of Rad51, p53 and overexpression of telomerase, telomeres became more fusogenic and dysfunctional (Tarsounas *et al.,* 2004.)

Other telomere associated proteins include the Ku heterodimer and WRN..

1.4.8 TERRA

Telomeric repeat-containing RNA (TERRA) is a large non-coding RNA that has been found in various mammalian species, Zebrafish and fungi via Northern blot analysis and RT-PCR by numerous research groups (Azzalin et al, 2007; Schoeftner and Blasco, 2008). Until the discovery of TERRA molecules, telomeres were thought to be transcriptionally silent (Azzalin et al., 2007). TERRA molecules range from 100bp to 9 kb, contain UUAGGG repeats and are located in nuclear fractions.

TERRA molecules are a key component of telomeric heterochromatin (Azzalin et al., 2007; Lingner et al., 2010). TERRA molecules have a role in the formation of heterochromatin and have been found to be located in close proximity to various known heterochromatic proteins, such as H3 trimethyl K9, and have been shown to associate with ORC1, which has also been found to have a role in histone formation (Deng et al., 2009).

TERRA molecules have also been shown to have a role in the structural stability of telomeres and telomerase regulation via interactions with shelterin components, such as TRF1 and TRF2 (Schoeftner and Blasco, 2008; Deng et al., 2009). Furthermore the depletion of TERRA molecules has been shown to trigger signs of telomere dysfunction by the presence of telomere dysfunction-induced foci (TIFs) and structural abnormalities (Deng et al., 2009).

TERRA molecule levels have been correlated with telomere length, for examples in ageing fibroblasts with short telomere lengths, TERRA levels decreased (Caslini et al., 2009). Furthermore, human cells have been shown to have lower levels of TERRA than laboratory mice which have longer telomeres (Schoeftner and Blasco 2008).

1.5: Telomere length dynamics

The enzyme telomerase catalyzes the addition of telomere repeats *de novo* (Yu and Blackburn, 1991). Telomerase counteracts the losses from the terminus that would arise as a consequence of the end replication problem (Blackburn, 1992). Telomerase is expressed in germ cells and stem cells, but not sufficiently in the majority of human somatic cells, resulting in gradual erosion of telomeric sequences with each cell division, at rates of between 50-200bp each cellular division (Counter et al., 1992). This gradual shortening of telomere sequence provides a barrier on the replicative lifespan of somatic cells as described by Leonard Hayflick (the 'Hayflick limit') (Hayflick and Moorehead, 1961). Telomeres eroded past a certain limit lose their function for protecting double stranded DNA and this loss of function triggers a DNA damage response inducing a p53 dependent G1/S cell cycle arrest known as 'replicative senescence' (Harley et al. 1990; Martens et al. 2000). Overexpression of telomerase in normal human somatic cells *in vitro* has demonstrated cells can bypass replicative senescence and can proliferate indefinitely thereby formally establishing the role of telomere erosion in the induction of replicative senescence (Bodnar et al. 1998; Vaziri and Benchimol 1998).

1.6: Telomere function in human cells

1.6.1: The end replication problem

The end replication problem explains the progressive loss of DNA sequence from the 5′ end of the lagging strand during replication (Olovnikov, 1971). In linear eukaryotes DNA replicates using an enzyme complex known as DNA polymerase (Kornberg et al., 1958). Replication is semi-conservative in that the two strands of DNA are separated and are used as a template for a new strand. DNA polymerase can initiate the replication of DNA in a 5′-3′ direction and requires the action of an RNA primer (Olovnikov, 1971). The leading strand (5′-3′) is continuously replicated by the polymerase; however the lagging strand DNA is synthesized in small pieces or ‘Okazaki fragments’ (Okazaki and Sakabe, 1966). At the extreme end of the DNA, the leading stand can be synthesised to the terminus, but the lagging strand cannot because an RNA primer is required to begin each piece of the lagging strand DNA, and thus at the end of the DNA there is nothing for this piece to attach to thus the last section of the lagging strand cannot be synthesized (Olovnikov, 1971). This truncation at the 5′ end of the lagging strand leads a gradual loss of DNA from the chromosome end (Olovnikov, 1971).

Figure 1.3: The end replication problem in human somatic cells

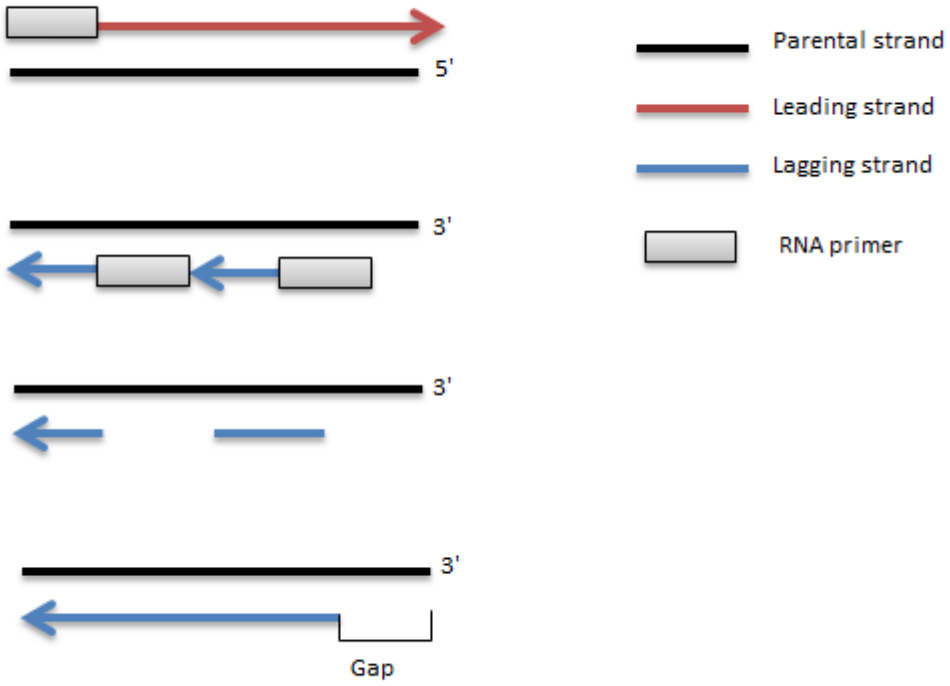


Figure 1.3: Simplified diagrammatic representation of the end replication problem and how DNA losses are generated with each replication (Adapted from'The disparity between human cell senescence in vitro and lifelong replication in vivo' by Harry Rubin Nature Biotechnology 20, 675 - 681 (2002) doi:10.1038/nbt0702-

1.7: Maintaining Chromosomal integrity

X-ray experiments involving both *Drosophila* and *Zea Mays* showed that chromosomal alterations caused by damage to DNA were excluded from the termini of chromosomes. This 'capping' effect distinguished the terminus of the chromosome from DNA double strand breaks induced by X irradiation (McClinktock *et al.,* 1938, Muller *et al.,*, 1939). This end-capping function of telomeres has been hypothesised to arise via the presence of t-loop lariat structures (mentioned previously) that were first visualised by electron microscopy (Griffith *et al.*, 1999) and also G-quadraplex structures. Telomeres have been shown to form G-quadraplex structures *in vitro* (Blackburn, 1991). G-quadraplexes are guanine rich with strings of four Gs, and are capable of forming a square planar structures or tetrads whereby two or more of these structures can stack on top of each other via Hoogsteen base-pairing to form a G-quadruplex (Burge et al., 2006). G-quadruplexes have been shown to decrease the activity of the enzyme telomerase, preventing the elongation of telomeres; in HeLa cells POT1 and TRF2 proteins binding affinity with to human telomeric DNA decreased when the single-stranded overhang is in the form of a quadruplex (Yanez et al., 2004).

1.8: Chromosome healing

In the 1930s when Barbara McClintock discovered the phenomenon of anaphase BFB cycles, one important question to answer was whether these cycles of breakage and fusion continue ad infinitum (McClintock, 1941). On analysis of inherited characteristics for plant colour in maize development, it was discovered that chromosomes can 'heal'. Introduction of a dicentric-generating chromosome (chromosome 9) that introduced breakage conferred loss of a dominant allele (McClintock, 1941). Observations showed that zygotes inheriting the dicentric-generating chromosome, after breakage post fertilization, generated newly formed

chromosomal ends behaving as natural termini and thus the BFB cycle was broken in the embryo. This addition of a new telomere to the site of DSB in telomeres was termed 'chromosome healing' (McClintock, 1941). Chromosome healing of double-stranded DNA breaks is considered to account for genetic conditions that arise as a consequence of terminal chromosomal deletions; these include alpha thalassemia (Wilkie, 1990; **Flint et al**., 1995).

1.9: TPE: The Telomere Position Effect

The telomere position effect (TPE) is a mechanism by which genes are reversibly silenced when they are located close to telomere repeat arrays, via modulation of chromatin structure. TPE has been studied in the budding yeast *S. cerevisiae,* however the role of the TPE mechanism in mammals is less understood [Reviewed by: Ottaviani et al., 2007]. TPE was discovered in yeast when studying chromatin structure in yeast telomeres (Gottschling *et al.,* 1990). TPE was observed in *S. cerevisiae* on the insertion of four reporter genes close to the telomere, these genes were silenced however the effects of this could be reversed (Gottschling et al., 1990; de Bruin et al., 2000). TPE has also been displayed in yeast strains expressing a uracil gene (URA3) (Sandell *et al.,* 1994). URA3 can be both selected for and against on different media allowing the observation of the 'variegated' or mosaic phenotype that is frequently seen in TPE; URA3 that was within 8bp of the telomere array grew on both media showing that gene switching can occur. Many proteins have been characterised in yeast that counteract TPE, such as the Sir proteins (2, 3 and 4) that interact with Rap1p, a direct telomere target (Aparicio et al. 1991; Laurenson and Rine, 1992; Kyrion et al., 1993). The human homologue hRap1p has also been shown to have a similar role in eliminating TPE silencing of gene expression. TPE has been postulated to respond to changes in telomere length to alter gene expression; the effect of TPE on silencing gene expression has thus been shown to be dependent upon both telomere length and distance of genes from the telomere

(Li et al., 2000; Arat and Griffith, 2012). In humans cells (HeLa) the use of a luciferase reporter gene in close proximity (~1.6 kb) to the telomere repeat array revealed a ten-fold decrease in luciferase expression compared with cells that had luciferase integrated in an internal site, implicating the TPE mechanism functioning to silence the reporter gene (Baur *et al.,* 2001). Furthermore overexpression of hTERT to elongate the telomeres resulted in a further silencing of gene expression, implicating telomere length as having an effect on TPE (Baur *et al.,* 2001). It has been hypothesised that gene expression could thus be pre-programmed via telomere lengths i.e. with ongoing cell divisions telomeres will shorten which could have an effect on the genes that are expressed implicating a role for telomere length in ageing and cancer via TPE related gene expression changes (Baur *et al.,* 2001).

1.10: Telomerase and telomere replication

1.10.1: Telomerase Structure and function

Telomerase is a ribonucleoprotein structure with an inbuilt reverse transcriptase. Telomerase was first discovered in the early 1980's by Elizabeth Blackburn and colleagues whilst studying the protozoa *Tetrahymena thermophilia* (Blackburn, 1981). Human telomerase consists of two molecules each of a telomerase RNA component (TERC), a reverse transcriptase (TERT) and dyskerin (DKC1), which binds the TERC component (Cohen et al., 2007). Human TERT polypeptide is a reverse transcriptase that folds inward with the TERC, a non-coding RNA. TERT has a holoenzyme structure that allows it to envelope the chromosome terminus to add single-stranded telomere repeats (Cohen et al., 2007).

In humans, the TERC RNA molecule includes the sequence 5′-UAACCCUAA-3′, which is complementary to the telomere repeat 5′-TTAGGG-3′, thus acts as a template for adding

telomere repeat arrays *de novo* to the chromosome end (Cohen et al., 2007) (Figure 1.2).

1.10.2: Telomerase expression in normal human cells and cancer cells

The majority of somatic cells in humans do not express telomerase and this leads to the subsequent loss of telomeric sequences at a rate of ~60–120 base pairs (bp) per cell division (Harley et al. 1990; Baird et al. 2003). Exceptions include B-and T-lymphocytes, endothelial cells and somatic stem cells are (Kim et al., 1994). Long-lived cell populations such as the germline and stem cells, which need a high self-renewal potential, can circumvent telomere-dependent replicative senescence and acquire unlimited proliferation capability by expressing telomerase (Kim et al. 1994). SV40-transformed cells have been shown *in vitro* to express telomerase on emergence from crisis, but not beforehand (Rhim et al., 1985; O Brien et al., 1986; Counter et al., 1992). Around 80–90% of malignant tumours overexpress telomerase and this high telomerase activity is associated with the high immortality of cancer cells and thus has a role in tumour development (Kim et al. 1994; Kolquist et al. 1998). A study by Hiyama *et al.* investigated the relative expression of telomerase in a range of different breast cancer subtypes. Findings from this indicated telomerase activity was detected in 68% of primary stage I breast cancers and 95% in the most advanced stage tumours. These observations point to a role for the activation of telomerase in later stages, which will result in stabilisation of telomere length and thus immortalisation of cells (Hiyama et al., 1996)

Figure 1.4: A simplified model for the action of Telomerase

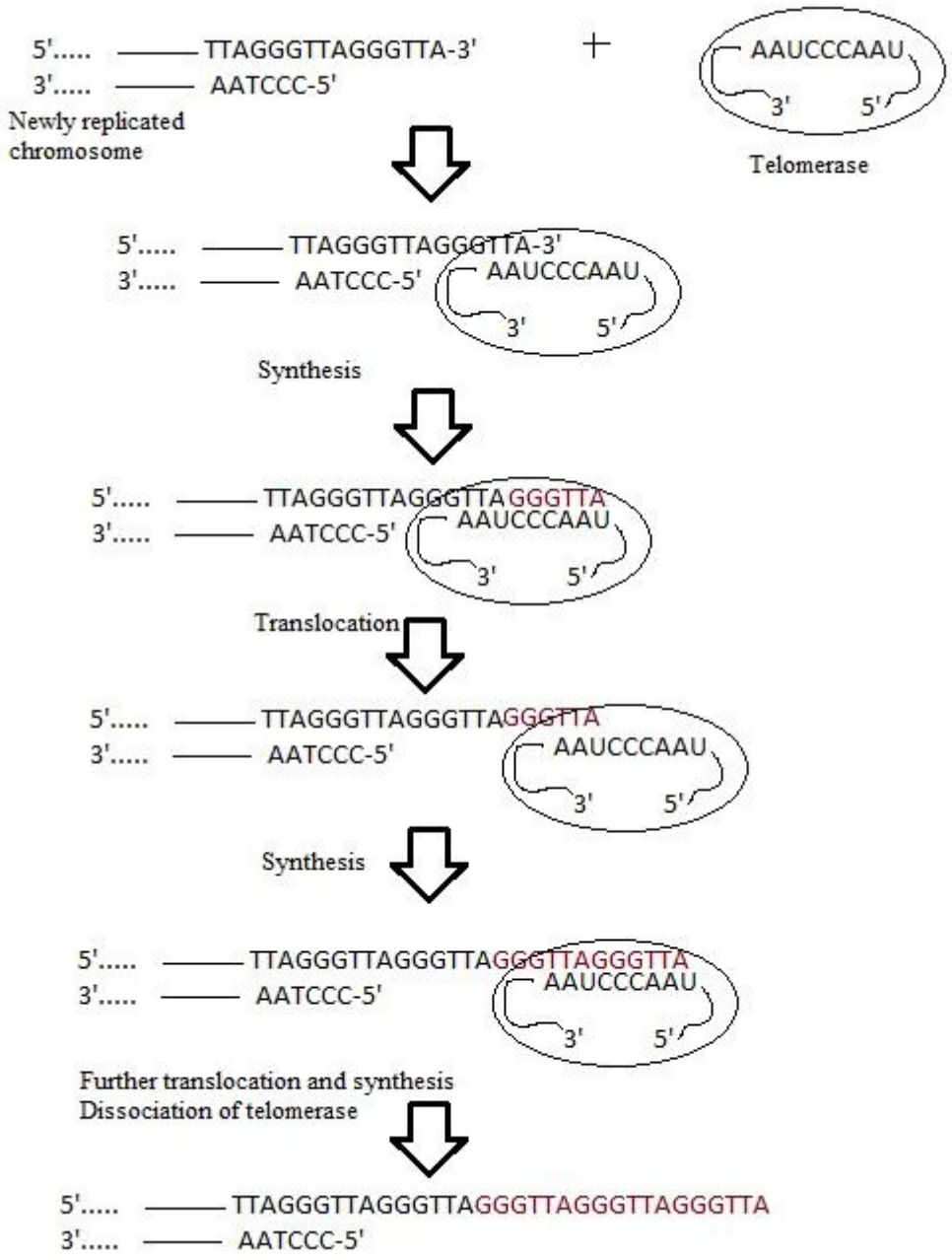


Figure 1.4: An illustration of how telomerase contains an RNA template against which new TTAGGG repeats can be synthesised onto the DNA terminus *de novo.* Adapted from Sweeney, G. (2008), "Telomeres; Genes and Development 2", Lecture, Cardiff University, School of Biosciences, unpublished.

1.11: Alternative Lengthening of Telomeres (ALT)

Subsets of SV40-immortalized cells have been found to be telomerase negative yet maintain telomere length (Bryan, 1995). It has since been discovered that around 10-15% of cancers show no evidence of telomerase activity (Bryan et al., 1997), thus cells that emerge from crisis can be immortalized through telomere elongation by an alternative pathway termed Alternative Lengthening of Telomeres (ALT). The mechanisms of ALT are largely still unknown; however it is thought to elongate telomeres via a recombinational mechanism (Dunham et al., 2000). Telomeres that maintain their lengths via ALT pathways have been observed to be extremely heterogeneous in length, ranging from 1 kb to over 20 kb (Rogan et al., 1995). Histologically presence of ALT is identifiable by ALT-associated promyelocytic leukemia (PML) protein nuclear bodies (APBs) that contain large amounts of extra-chromosomal telomeric DNA, PML protein and other proteins involved in telomere binding, DNA replication, and recombination (Guo, A et al. 2000, Subhawong, et al. 2009). The regions of extrachromosomal telomeric DNA may be presented as double stranded telomere circles (t-circles) (Cesare and Griffith, 2004). APBs may be the site of telomere lengthening, or they may hold the protein complexes required for elongation however the exact function remains unclassified (Subhawong, et al. 2009).

1.12: DNA Damage Repair Mechanisms

A number of mechanisms are in place to ensure the faithful replication of DNA, guaranteeing the correct genetic information is inherited by daughter cells (Lodish et al., 2004). There are however thousands of potential damageing events that can interfere with the stability of the DNA such as a break or lesion (Freidburg et al., 1995). It is important to repair damage inflicted upon DNA so that replication errors do not occur and mutations are kept to a minimum to maintain stability of the genome (figure 1.5). Damaged DNA that has faulty DNA repair mechanisms can bring about genome instability, which can have an impact on ageing due to loss of functionality of genes and resulting apoptosis/senescence (Hoeijmakers, 2001; Gottshling et al., 2003; Aguilera and Gonzalez, 2008). Genome instability can lead to unregulated cell division via an accumulation of DNA damage and the inability to repair faults in the DNA, thus potentially leading to cancer (Hoeijmakers, 2001; Campisi and **d'Adda di Fagagna** , 2007; Aguilera and Gonzalez, 2008).

Both endogenous and exogenous threats can interfere with the integrity of DNA (Aguilera and Gonzalez, 2008). Endogenous factors causing disruption include oxidation, hydrolysis, alkylation and the mismatch of bases. Exogenous damage can occur from exposure to chemical agents such as hydrogen peroxide, ionizing radiation (IR) and ultraviolet light (UV-A and UV-B) (Lodish et al., 2000; Marnett and Plastaras 2001). Damage to DNA, such as a DNA break, alters the structure of the DNA helix detectable by the cell. Once damage is detected or localised, DNA repair machinery can be recruited at or near the site of damage (Lodish et al., 2004).

Single stranded DNA breaks occurs more frequently than double stranded breaks (Bradley et al., 1979). Single stranded DNA damage occurs when just one strand of the double helix is damaged and thus repair of single stranded breaks (SSB) makes use of the unaffected strand as a template to guide the repair (Caldecot, 2008). SSB repair involves three main excision

repair pathways that aim to remove the damaged nucleotide section and replace it with complementary sequence to the unaffected strand (Caldecot, 2008). Mismatch repair (MMR) involves the correction of mispaired nucleotides; base excision repair (BER) replaces single bases and nucleotide excision repair (NER) repairs bulky lesions such as pyrimidine dimers [Reviewed by Caldecot, 2008).

Double stranded breaks (DSB) can be especially damageing because if not repaired, they can lead to large scale genomic rearrangements (Griffith et al., 1999). There are two fundamentally different mechanisms of DSB repair: homologous recombination and non-homologous end-joining.

Figure 1.5: The importance of DNA damage repair

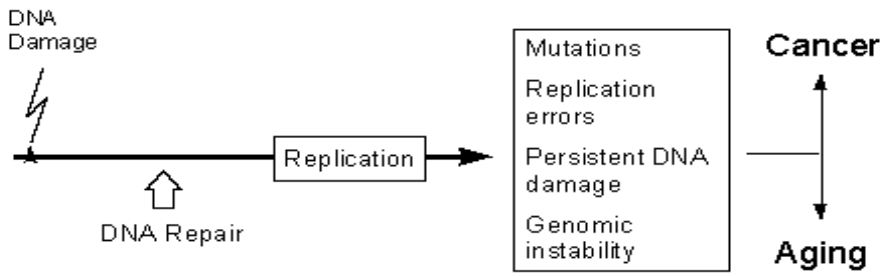


Figure 1.5: In the absence of a functional DNA damage repair during replication, resulting mutations, replication errors, DNA damage and genomic instability can arise. These effects can lead to both ageing and cancer (Adapted from Day, 2007).

1.14:DSB repair mechanisms

Non-homologous end joining (NHEJ) and homologous recombination are two key

mechanisms that repair double-strand breaks (DSB) in DNA (Wyman and Kanaar, 2006). NHEJ involves the direct ligation of ends without a template contrasting with homologous recombination which requires a homologous sequence template to guide repair (Wyman et al, 2004; O' Driscoll and Jeggo, 2006).

1.13.1 Homologous Recombination (HR)

Homologous recombination (HR) involves the exchange of nucleotides between similar chromatids (Lodish et al., 2000; Wyman et al, 2004). HR typically occurs in either the S phase or G2 phase of the cell cycle, the availability of sister chromatids for homologous templates is important so repair of DSB via this mechanism occurs before the mitotic 'M' phase of the cell cycle (Alberts et al., 2008). A family of cyclin dependent kinases (CDKs) regulate HR in eukaryotic organisms (Shrivastav et al., 2008). The HR model of DNA damage repair mechanisms has several key steps (Szostak et al., 1983). Initiation of HR involves the binding of the MRN complex either side of the ds DNA break (Szostak et al., 1983; Lodish et al., 2000). Processing occurs whereby sections of DNA around the 5′ termini of the breaks undergo nucleolytic resection giving rise to single stranded tails with free 3′-OH ends (Szostak et al., 1983). Strand invasion occurs whereby homologous regions between the DNA break and the sister chromatid template are recognised and form a displacement or D-loop intermediate (Szostak et al., 1983). DNA polymerase synthesises a new strand extending from the 3′ end of the invading strand (Szostak et al., 1983). The second DSB is sequestered to form another D-loop. Strands cross over forming a double Holliday junction. Gap filling DNA synthesis and ligation then completes the new DNA strands (Szostak et al., 1983).

1.13.2 Non-homologous end joining (NHEJ)

The NHEJ mechanism starts with recognition of the DSB by the action of a Ku protein

(Walker et al., 2001). Ku protein is a heterodimer consisting of Ku70 and Ku80 polypeptides and binds to the end of DNA thought to act as a molecular scaffold onto which other proteins and end-processing enzymes can attach (Walker et al., 2001). The binding of Ku70/80 attracts and sequesters the catalytic subunits of DNA-PKcs, forming a complex (Yoo and Dynan, 1999). This interaction of the catalytic subunit DNA-PKcs with Ku activates the kinase function of the complex and thus initiates terminal processing events that involve autophosphorylation, phosphorylation of several other proteins, alongside polymerase and nuclease interactions (Lee-Miller and Meek, 2003). Ligation finalises the process by the acquisition of the DNA ligase IV/XRCC4 complex to stabilise and ligate the breaks (Lee-Miller and Meek, 2003).

1.13.3 Alterative non-homologous end joining: Microhomology-Mediated End Joining (MMEJ)

An error prone mechanism of end joining has been characterised recently called microhomology-mediated end joining (MMEJ), whereby 5-25bp of homologous sequence is used to guide the alignment of broken DNA strands prior to ligation. MMEJ has been considered a 'back up' when NHEJ fails and has been associated with chromosomal alterations such as translocations and deletions, however recent studies have shown MMEJ to be the mechanism at play regardless of the presence of functional classical NHEJ components (Lee-Theilen et al., 2011; Symington and Gaultier, 2011). During V(D)J recombination in the developing B-cells, MMEJ has been proposed as an alternative pathway of recombination in the absence of DNA ligase IV or XRCC4 (Boboila et al., 2010). Mice lacking functional p53 and Ku subunits (or DNA ligase IV) have been shown to develop cancers such as lymphomas and accumulate non-reciprocal translocations with evidence of microhomologies at the breakpoints (Maser et al. 2007). Recent studies by Letsolo *et al.* have shown evidence of

MMEJ as a process of end repair following large scale deletion events occurring when dysfunctional telomeres undergo fusion events in telomerase negative MRC5 clonal fibroblasts expressing human papillomavirus oncoproteins E6 E7 (Letsolo *et al.*, 2010).

Evidence in species such as *Xenopus* and calf (thymus) has shown a Ku independent mechanism of end-joining with alternative characteristics such as large deletions and regions of microhomology. It has also been shown that it is not only the lack of Ku that can cause an alternative end-joining mechanism, but the loss of other NHEJ components such as ligase IV and XRCC4 (Mason et al., 1996; Gottlich et al., 1998).

Repair of DSBs arising from ionising radiation and radiomimetric agents in bladder cancer tumours have been observed to preferentially undergo an error prone alternative end-joining mechanism mediated by microhomologies, rather than the classical NHEJ pathway (Bentley et al., 2004). Ku independent microhomology mediated end-joining has now been observed in many different species including fungi (*S. cerevisiae)* (Maj et al., 2003).

1.14.1: Telomere dysfunction and telomere fusion

The successive shortening of human telomeres can induce replicative senescence, which is the arrest of cellular division in response to DNA damage via the p53 and pRb (cell-cycle checkpoint) pathways (Campisi, 1997). This is also known as the M1 stage and is a natural tumour suppressor mechanism providing a barrier to unlimited cell division (Wright and Shay, 1992). In the absence of a functional DNA damage response, cells bypass M1 (senescence) and telomeres erode further until they undergo crisis (M2 stage) (Wright and Shay, 1992). M2 is characterised by telomeres that have lost their ability to cap the

chromosomal ends, resulting in end-fusions (Wright and Shay, 1992; Counter et al., 1992). This can result in potentially oncogenic genomic rearrangements such as NRTs, aneuploidy and gene amplification that are consistent with telomere dysfunction (Counter et al., 1992; Capper et al, 2007; Lin et al., 2010). Genomic instability has been correlated with the progression to malignancy in many types of cancer (Negrini et al., 2010; **Shen , 2011**)). Non-reciprocal translocations (NRTs), a type of genomic rearrangement, are thought to be key mutational events that can drive many types of malignancy (Mitelman et al., 1997). Telomere fusion events have been detected alongside NRTs in the progression of Chronic Lymphocytic Leukaemia (CLL) (Lin et al., 2010)

In normal cells/tissues that exhibit otherwise long a stable telomeres, large-scale deletion events that can create telomeres that are capable of fusion, this can occur in both in the presence or absence of telomerase (Baird et al. 2003, 2006; Britt-Compton et al. 2006). A fusogenic threshold (<42 pure TTAGGG repeats) has been determined using single molecule based techniques in telomerase-negative human cells *in vitro* below which telomeres are deemed a 'critical' length and are subjected to fusions (Capper et al., 2007; Letsolo et al., 2010).

Telomere fusion events involving regions close to known fragile sites have been identified, for example fusions involving telomeres and the non-telomeric loci, 2q14, 8q24.3, 7p21.1 and Xp22.1, all of which have been documented as being in the same cytogenetic location at fragile sites (Letsolo et al., 2010; Debacker, and Kooy, 2007). A number of fragile sites are thus associated with human disease and cancer (Debacker and Kooy, 2007), for example at FRA3B is the gene FHIT (fragile histadine triad protein), and encodes an enzyme that acts as a HER2 tumour suppressor (Bianchi et al., 2007). Large scale genomic rearrangements could potentially be generated by the fusion of dysfunctional telomeres with a fragile site, facilitating the progression to malignancy.

1.14.2 The mechanisms of telomere fusion:

The mechanistic basis for telomere fusion is still unclear. One study showed that inhibiting TRF2 resulted in fusion that was dependent on factors involved in classical-NHEJ (non-homologous end joining) (Smogorzewska et al. 2002; Celli and de Lange 2005). Telomeres have a role in protecting eukaryotic chromosomes from fusions by that action of NHEJ via proteins bound to telomeric loci. In mammals TRF2, involved in the t-loop structure, inhibits aberrant NHEJ activity at the chromosomal terminus and prevents telomere fusion (Chong et al. 1995; Broccoli et al. 1997; Baumann and Cech 2001). Fusions in normal cells are thus a rare occurrence, but short dysfunctional telomeres that have lost their protective 'cap' are more readily fusogenic (Capper et al., 2007).

Telomere fusions can lead to the formation of dicentric chromosomes which go on to initiate cycles of anaphase breakage-fusion-bridging as first characterised by Barbara McClintock (McClintock, 1941). Evidence of such dicentric bridging has been observed extensively, elucidating roles for TRF2 in fusion prevention and thus telomere stability (Van Steensel et al., 1998; Smogorzewska et al. 2002; Celli and de Lange 2005). Van Steensel and colleagues noticed an increase in the presence of anaphase bridges when inhibiting TRF2 in human cells and Denchi et al. observed extensive telomere fusions when TRF2 was deleted from mouse liver cells (Van Steensel et al, 1998; Denchi et al, 2006). Parp1 has also been shown to have a role in protecting telomeres from fusions via interactions with TRF2. Parp1 deficient murine cells show an increase in telomere fusions when DNA damage is induced (Gomez et al., 2006).

In vitro cell assays of double strand break (DSB) repair have revealed non homologous end-joining (NHEJ) as a key mechanism by which ends are joined accurately, efficiently and

without any loss of nucleotides (Labhart et al., 1999). Double strand break repair pathways involving NHEJ are thought to account for most chromosome fusions (Smogorzewska et al. 2002; Celli and de Lange 2005). There are, however, different types of NHEJ mechanism. Classical NHEJ requires a Ku heterodimer to recognise a DNA ds break and act as a molecular scaffold onto which end-processing enzymes can facilitate repair of the broken ends (Labhart et al., 1999). Alternative NHEJ mechanisms in the absence of Ku and ligase IV utilise regions of microhomology to facilitate repair of broken ends and this mechanism has been observed to be associated with large scale deletions and chromosomal rearrangements (Maser et al. 2007; Lee-Theilen et al., 2011; Symington and Gaultier, 2011).

1.15: The Role of Telomeres in Genomic Instability

Faithful replication of DNA is constantly challenged (Lodish et al., 2004). Numerous processes involved in cell division are stringently controlled to maintain the integrity of the genome (Lodish et al., 2004). Genomic instability can drive evolution via genetic variation (Maizels et al., 2005; Aguilera and Gonzalez, 2008); however mutations and rearrangements associated with genomic instabilities are frequently associated with inherited diseases, ageing and cancer [Reviewed by O'Sullivan and Karlseder, 2010; Negrini et al 2010]. There are many types of genomic instability and ways in which aberrant changes arise.

Chromosomal changes or aneuploidy, involve changes in chromosome number via attrition or loss altering the karyotype (Sen, 2000). Sister chromatid exchange (SCE) can be a sign of DNA damage repair because it is a process that aims to resolve double stranded breaks that occur though replication errors (Cortes-Ledesma and Aguilera, 2006). Unequal sister chromatid exchange (SCE) is the joining of ends between non-homologous DNA strands generated by DNA breaks and because of this SCE can result in large scale chromosomal rearrangements such as duplication or deletions (Aguilera and Gonzalez, 2008).

A large number of genomic alterations occur during the S-phase, via replication errors such

as replication slippage (Viguira et al., 2001). Replication slippage gives rise to the deletion or insertion of repeat arrays or 'microsatellite polymorphisms' due to the misalignment of DNA strands during replication (Viguira et al., 2001; Aguilera and Gonzalez, 2008). Replication slippage thus causes DNA sequence gains or losses. This is why replication checkpoints are important. In humans, ataxia telangiectasia mutated (ATM), ataxia telangiectasia and RAD3 related (ATR) trigger an S-phase checkpoint response to reduce replication errors (Abraham, 2001). ATR is a transducer kinase that targets stalled replication forks (RF) and accumulation of single stranded DNA (Cortez et al. 2001; Rouse and Jackson 2002; Zou and Elledge 2003). ATM targets DSBs and is recruited via an MRN complex (Lavin 2007). Both kinases phosphorylate effector kinases CHK1 and CHK2 (Kim *et al.*,1999; Matsuoka *et al.* 2000). If this response is disrupted DNA breaks accumulate (Abraham, 2001).

Other mutational processes can cause genomic instability such as deletions, inversions and translocations. These aberrant changes can arise in many ways, such as through impaired homologous recombination, non-homologous end joining and the error prone microhomology-mediated end joining [Reviewed by Aguilera and Gonzalez 2008].

Another way in which genomic instabilities can arise is through telomere fusions resulting from telomere dysfunction (Lin et al., 2010). Gross chromosomal rearrangements can arise via cycles of B-F-B leading to unfavourable alterations (such as NRTs) that potentially drive oncogenic rearrangements (Artandi et al., 2000; Murnane et al., 2004). Non-reciprocal translocations (NRTs), a type of genomic rearrangement, are thought to be key mutational events that can drive many types of malignancy (Mitelman et al., 1997, Artandi et al., 2000). Evidence has shown that NRTs can arise from telomere fusion events in the progression of Chronic Lymphocytic Leukaemia (CLL) (Lin et al., 2010).

Telomere length-related genome instability has been extensively modelled in mice (Rudolphe et al., 2001; Artandi and dePinho, 2010). Successive generations of mice with a loss of

mTERC components in APC^{min} models (APC mutations cause multiple intestinal neoplasia) were used to relate telomere dysfunction and genomic instability (Rudolphe et al., 2001). Findings revealed an increase in detectable anaphase bridges (by *FISH*), at the adenoma to carcinoma transition, which was postulated to implicate telomere dysfunction related genomic instability as having an early stage oncogenic effect (Rudolphe et al., 2001).

In prostate cancer, coupling telomere studies (by TRF and *FISH*) with cytogenetic analysis to assess genomic instability have shown correlations between telomere shortening and genomic instability in the early stages of prostate cancer (Meeker et al., 2002). Telomere loss and genomic instability have also been modelled in early stages of human prostate cancer using p53/PTEN knockout mice with successive loss of mTERT (Artandi and dePinho, 2010, Ding et al., 2012). Activation of telomerase following the induction of telomere dysfunction allowed an accumulation of genomic instabilities consistent with the idea that in early stages of prostate cancer telomere dysfunction can lead to genomic instability and the reintroduction of telomerase can maintain cells with instabilities leading to cancer progression (Ding et al., 2012).

Telomere lengths in peripheral blood measured by Q-PCR in both sporadic and hereditary ovarian cancer patients revealed that short telomeres were associated with earlier age of onset in both subtypes (correlation r=0.32; P=3.1 x 10^{-12}) (Martinez-Delgado et al., 2012). Short telomeres were postulated to thus be contributing to genomic instability hence accelerating tumourigenesis (Martinez-Delgado et al., 2012).

Direct telomere analysis using high resolution single molecule techniques (STELA; Baird et al., 2004) has revealed evidence of short telomeres that are capable of telomere fusion events coupled with evidence of large scale genomic instability via array-CGH in CLL patient samples (Lin et al. 2010).

1.16 Telomere dependent replicative senescence

Replicative senescence is a mechanism that has roles in both cancer and ageing [Reviewed by Shay and Wright, 2005).

Replicative senescence is an essential mechanism employed by somatic cells and is often excluded from tumour cells (Campisi, 1997). Replicative senescence impedes unlimited cell proliferation (Campisi, 1997). Senescence can be telomere dependent or telomere independent.

Telomere dependent replicative senescence arises when telomeres reach threshold lengths acting as a sensor for the number of cell divisions that have taken place, termed the 'Hayflick limit' (Hayflick and Moorehead, 1961; (Harley et al., 1990).). Replicative senescence is brought about by DNA damage response (DDR) pathways via tumour suppressor proteins p53 and pRb. P53 and pRB determine whether or not a cell goes into senescence. Telomeres that have lost their ability to cap chromosomal termini and thus initiate a DNA damage checkpoint response by activation of the ATM/p53 pathway (Blackburn, 2000**, Artandi et al**., 2005). ATM (Ataxia telangiectacsia mutated) activates CHK2, an effector kinase, by phosphorylation which in turn inhibits CDC25 (Chehab et al., 2000). CDC25 is a phosphatase that regulates entry into and progression through various phases of the cell cycle (Gould et al., 1991). This thus halts progression through the cell cycle. P53 is a tumour suppressor protein that recognises telomere dysfunction as a DNA break and activates a downstream process involving increasing the expression of p21 and inhibiting the phosphorylation of pRB [Reviewed by Artandi et al., 2005]. P16 is a positive regulator of the pRB protein and has a role in controlling the cell cycle via cyclin dependent kinases (CDKs) (Kulju and Lehman

1995; Alcorta et al. 1996; Hara et al. 1996; Reznikoff et al. 1996; Herbig *et al.*, 2004). Senescent human fibroblasts have been shown to overexpress p21 and p16, signalling proteins that inhibit cyclin dependent kinases upstream of pRb (van Steensel et al. 1998). Ectopic expression of telomerase subunit hTERT causes telomere elongation and thus senescence can be overcome (Bodnar et al. 1998).

Figure 1.7: Pathways connecting telomeres and p53 in senescence, apoptosis, and cancer.

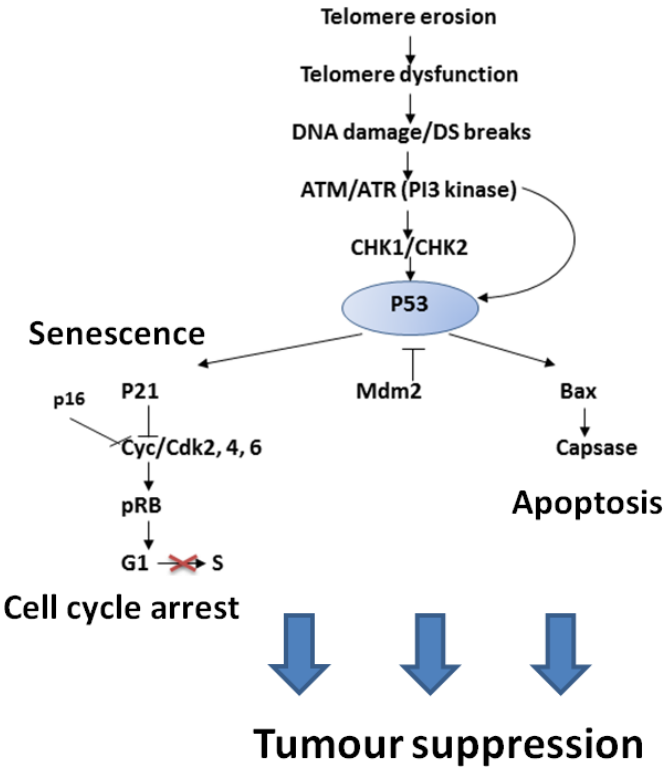


Figure 1.7: A diagrammatic representation of how telomeres can affect p53 and its related pathways. Telomere erosion can lead to cycles of B-F-B, and the resulting DS breaks that can arise lead to the activation of ATM/ATR. Downstream targets of p53: P21 prevents progression through the cell cycle; Mdm2 negatively regulates p53 and BAX competitively inhibits bcl2 expression which results in the activation of the capsase cascade leading to apoptosis. All these processes enable tumour suppression (adapted from Artandi et al., 2005)

1.17 Telomere independent senescence

Telomere independent senescence has been observed via the overexpression of cellular and viral oncogenes such as SV40, E6 and E7. Evidence has shown, for example, that expression of the SV40 large T antigen dysregulates the function of p53 and pRb genes resulting in cells that can bypass replicative senescence (Shay et al. 1991) and thus are immortalised. Cells that can bypass senescence extend their natural replicative lifespan and thus have the ability to accumulate larger numbers of mutations and DNA damages that can lead to the initiation of neoplasia. In mouse models, restoration of p53 function causes the reduction of tumours such as sarcomas via a senescent-like phenotype involving cell cycle arrest (Ventura, et al. 2007).

Another type of telomere independent senescence can occur *in vitro* as a result of unfavourable culture conditions such as oxidative stress (Chen et al. 1995) or hypoxia, which is also referred to as STASIS or Stress or Aberrant Signalling Induced Senescence, premature senescence or culture shock.

1.18: Ageing

Replicative senescence has been suggested to have evolved as a protective mechanism against unwanted cell proliferation, for example due to oncogene overexpression (Campisi, 2000). The successive shortening of telomeres triggers senescence as a natural function of age (Harley et al., 1990). Phenotypically cells that are senescent grow larger, with a flatter shape and express senescence-associated β-galactosidase, a marker of cellular senescence (Bayreuther et al., 1988; Dimri et al., 1995).

Replicative senescence is considered to underlie, at least in part, the ageing process in humans. Ageing can be characterised by the deterioration of tissue function and has been shown in many species to be related to the accumulation of senescent cells (Röhme, 1981). Senescent cells can secrete proinflammatory cytokines and other factors that disrupt the tissue

microenvironment and thus may contribute to disruption of cell and tissue function (Campisi et al., 2000; Shay and Wright, 2000). Some studies show that senescent cells contribute to age-related pathologies such as osteoarthritis (Martin and Buckwalter, 2002; Price et al., 2002) and to skin aging (Giacomoni et al., 2000). Cells from humans and primates that are relatively long-lived species have been shown to senesce after more cell divisions than short lived species (Hayflick et al., 1977). In telomerase deficient mouse models, knockout of TERC caused the progressive anticipation of pathologies that are associated with loss of telomeric repeats and ageing such as heart failure and decreased tissue regeneration (Blasco *et al.*, 1997 and 2005). In WRN protein mouse knockout models symptoms of premature ageing were correlated with short telomere lengths (Chang et al., 2004; Du et al., 2004). Humans with hereditary premature ageing syndromes such as dyskeratosis congenita (DKC) have been shown to display genetic anticipation in relation to a decline in telomere lengths and this decline in telomere length is associated with earlier onset of age related phenotypes such as grey hair, tissue deterioration and cancer disposition (Vulliamy et al., 2008).

1.19.1 Telomere length dynamics affected in genetic disease: Dyskeratosis Congenita

Dyskeratosis congenita is a genetic disease whose pathogenesis is directly related to telomere length maintenance (Vulliamy et al., 2001). Mutations in DKC1 andTERC are commonly found in patients with dyskeratosis congenita (Vulliamy et al., 2001). DKC1 and TERC are important components of telomerase, the enzyme that catalyses the addition of telomere repeats in order to maintain telomere lengths. DKC1 encodes the protein ‘dyskerin’ which is involved in attrition of TERC, cells lacking this protein thus have reduced levels of TERC RNA leading to the inability of the enzyme telomerase acting to elongate the telomere end. The disease also shows autosomal dominance due to haploinsufficiency i.e. the sufferer

produces half the normal amount of telomerase (Vulliamy et al., 2004) andanticipation resulting in progeny of affected parents with a worse prognosis than preceding generations (Vulliamy et al., 2004).

1.19.2 Other conditions: Werner's syndrome Unusually short telomeres are also seen in some inherited disorders associated with premature ageing such as Werner's syndrome. Werner's syndrome is also referred to as 'adult progeria' and is a rare autosomal recessive disease linked to mutations in the WRN gene found on chromosome 8, and with a high incidence of cancer (Epstein et al., 1966). Telomere dysfunction and genome instability has been observed in patients with Werner's and has been hypothesised as having a causal role in the premature ageing that is a hallmark of this disease. Mice lacking mTERC, the telomerase component showed similar pathologies to mice that lack the WRN gene (Chang et al., 2004). As a single dysfunctional telomere was found to be sufficient to limit cell survival *in vivo* (Hemann et al., 2001), it was thus considered that the presence of a few dysfunctional telomeres may be sufficient to cause the dysfunctional phenotypes in WRN (Chang and Multani, 2007). However the utilisation of single molecule PCR strategies revealed that telomeres in Werner's syndrome cells eroded at rates similar to those of normal controls (Baird et al., 2004), suggesting that the ageing phenotype of these cells is due to a stress-induced, p38-mediated growth arrest that is independent of telomere erosion (Baird et al., 2004). Studies indicate that an absence of WRN may indirectly effect telomere dynamics via cell growth kinetics (Wyllie et al., 2004, Baird et al., 2004).

1.20: Cancer

Cancer can be defined as the disease caused by an uncontrolled division of abnormal cells in a part of the body and also a malignant growth or tumour resulting from such a division of cells. Cancer cells have many different properties to distinguish them from normal cells of the human body. They have the ability to divide indefinitely, can metastasise to invade neighbouring tissues, fail to complete normal differentiation, can stimulate angiogenesis by releasing growth factors and do not require external signals to promote their growth. Cancer can arise from a number of different factors; carcinogens in the form of mutagens and non-mutagens, oncogenic viruses such as the Human Papilloma Virus (HPV) and also genetics i.e. inherited mutation in oncogenic genes such as the BRCA-1 gene in breast cancer.

1.21: Human Breast Cancer

Breast Cancer is a multifactorial disease that affects 13.4% of women and 6% of men and women combined worldwide (World Cancer Report, 2009). It is the most common malignant disease affecting women in the western world. Historically breast cancer is one of the earliest detectable carcinomas due to the visible phenotypic observations of tumourous lumps. Indeed possible breast cancer tumour diagnoses dates back to 1600 BC in Egypt (Uroskie and Colen, 2004).

Human breast cancer is most commonly characterised by tumours present in either the lobules surrounding the mammary ducts or the ducts themselves (Stewart, 1941; Foote and Stewart, 1946), but other rarer forms of breast cancer include Paget's disease of the nipple and inflammatory breast cancer affecting lymph vessels in the epidermis of the breast. Ductal and lobular carcinoma may begin with what is termed usual ductal or lobular hyperplasia (UDH/ULH), which is a benign overgrowth of cells lining the ducts or lobules with a pattern that is similar to normal cell histology. Atypical ductal/lobular hyperplasia (ADH/ALH) is a

slightly more abnormal class of benign growth in a confined region that gives the patient a higher (3½ to 5 times) risk of developing breast cancer. Ductal or lobular carcinoma *in situ* (DCIS/DLIS) is a neoplastic tumour still confined to its region, a pre-invasive cancerous stage (Allred et al, 2001; Richie and Swanson, 2003). Invasive ductal or lobular carcinoma occurs when the carcinoma *in situ* metastasises beyond basement membranes to invade surrounding parenchyma. Invasive ductal carcinoma (IDC) is more common than invasive lobular carcinoma (ILC) occurring in around 75-80% mammary carcinomas compared to the 10-15% comprised by ILC. ILC has been shown to have a slightly worse prognosis when compared with IDC (**Klem**i et al., 1997).

1.22: Epidemiology of Breast Cancer and risk factors associated with the disease

Epidemiology is a useful tool for identifying risk factors that have contributed to the improvement in detection and clinical care of the disease. Between 1971 and 1975, 5 year disease free survival averaged at about 52%. Today there is an 80% survival rate over 5 years, signifying a huge improvement. Regardless of this, breast cancer is still the most commonly diagnosed carcinoma of the western world, the lifetime risk of a female being diagnosed in the UK alone is 1 in 8 and in the last 10 years, breast cancer incidence rates in the UK have increased by 6% (Cancer research UK statistics).

There are multiple risk factors for developing breast cancer. Age is the largest contributor to probability of developing breast cancer, after gender, with people over 50 years of age having an 81% more risk than those below 50. This could be due to an accumulation of DNA damage with successive cell divisions in ageing tissues increasing the likelihood of oncogenic mutations.

Reproductive history is another factor to be considered to be strongly associated with breast cancer risk due to the circulation of the hormones oestrogen and progesterone cancer (Willett et al, 1989). Incidence of breast cancer has been shown to be affected by a number of different documented factors such as: age at menarche with a low age correlated with an increased risk; parity; late menopause; the use of oral contraceptives; age of woman for the birth of a first child [Reviewed by Bernstein, 2002]. Oestrogen and progesterone have been shown to increase mammary cell proliferation, leading to an increased risk of the accumulation of deleterious mutations via errors in DNA replication. Oestrogen also has been shown to encourage growth of the ductal systems in the breast and progesterone stimulates alveolar formations.

Prolactin is a peptide hormone that is has a key role in mammary gland development; studies in mice have shown a correlation between high levels of prolactin and an increase in mammary tumours (**Biswas and Vonderhaar**, 1987 and Welsch et al., 1977).

The inheritance of mutated breast cancer susceptibility genes (BRCA1&2 and p53) increase the risk of developing breast cancer by around 65% and also give rise to earlier onset of breast cancer compared with sporadic breast cancers. However, hereditary breast cancers only account for 5-10% of breast cancers.

1.23: The Methods for Classification of Breast Cancer

Breast cancer is not a single disease; it is histologically, clinically, genetically and

biochemically heterogeneous.

1.24. Pathological assessment of tumour type

By observing breast cancer tissue phenotypically under a microscope, the nature of the tumour relating to the position and malignancy can be characterised. The most common classes of breast cancer relate to spread so are invasive or *in situ*; location of the cancer means the tumour is termed either lobular or ductal. Invasive ductal carcinoma is one of the more common subtypes of breast tumour accounting for over 50% of breast tumours; however the World Health Organisation (WHO) recommends a large list of breast tumour pathological types for both benign and malignant tumours.

1.24.2: Histological grading of Breast Cancer

The Scarff-Bloom-Richardson (SBR) system of grading breast tumours microscopically was developed in 1957 and has since been adapted for higher quantitative accuracy by Ellis and Elston, and this adaptation on the SBR histological grading system is now referred to as the Nottingham Prognostic Index or Nottingham Grade. The grading is assessed from a tumour sample taken from a breast biopsy, lumpectomy or mastectomy. Phenotypic characteristics are scored from 1-3 relative to severity in terms of prognosis (1 being low risk, 3 confers a high risk). The factors that are considered include frequency of cell mitosis (mitotic index), percentage of tumour that is constituted of tubular structures and changes in cell size and uniformity. The scores are then added together for each category to give a final numerical value that allows the tumour type to be classed into a grade of I, II or III.

1.24.3: TNM Classification of Malignant Tumours (TNM)

The TNM system of classification is employed for all solid tumours and was designed in 1943 (Denoix et al., 1946). 'T' relates to the size of the tumour and is scored from 1-4; 'N' stands for nodes and relates to the spread of the tumour cells to lymph nodes and is scored 0-3; 'M' stands for metastasis and 0=no metastasis, 1= metastasis beyond regional lymph nodes. A lower score indicates a lower grade of tumour type and thus the better the expected prognostic outcome. Grading assessed by TNM stageing categorises tumours into grades 1-4, grade 4 tumours tend to have spread and tend to be inoperable thus indicate a poor prognostic outcome.

(see Table 1. 4).

Table 1.4

Stage	Description
0	A pre-malignant disease or marker (DCIS/ LCIS)
1-3	'Early' cancer with a good prognosis.
4	'Advanced' and/or 'metastatic' cancer with a bad prognosis.

Table 1.4: Stage that is categorised by the TMN system alongside prognostic description of tumour

1.24.4: Receptor status

1.24.5 Hormone Receptor (HR) Positive Breast Cancer

Breast cancer can also be classified by receptor status, which is dependent upon the presence of cell or hormone receptors (HR). Breast cancer cells may be positive for oestrogen receptor (ER) or progesterone receptor (PR)

Around 70% of all breast cancers are ER positive and of these around 60% are also PR positive. (Masood, 1992) ER+ (oestrogen receptor positive) and PR+ (progesterone receptor positive) breast cancer cells depend on presence of each hormone for their growth, respectively, thus drugs targeting oestrogen and progesterone receptors, such as oestrogen receptor antagonist Tamoxifen, can be used to block the hormone binding result in a block in cell proliferation and therefore a better prognosis.

Around 70% of breast cancers overexpress oestrogen receptors (Masood, 1992). There are two subtypes of oestrogen receptor: ERα and ERβ, which are different in size (ERβ is smallest) and encoded by different genes at different chromosomal loci (ESR1 on chromosome 14 and ESR2 on chromosome 6 respectively). However, both isoforms share homology in binding domains (Anderson, 2002). Tamoxifen treatment in high risk women has shown a reduction in breast cancer incidence (Fisher et al, 1998).

Progesterone receptor (PR) proteins exist as two isoforms: A and B. The PR isoform type A has been shown to repress the action of the B-isoform (of PR) and ER. The balance of the presence of these two isoforms in breast cancer has been studied for prognosis and studies have suggested that tumours with a higher proportion of PR-A are more likely to relapse and have a reduced response to hormone therapy such as Tamoxifen (Torsten et al, 2004; Badve and Nakshatri, 2008).

1.24.6 HER2 positive Breast cancer: over-expression of the *ERBB2* gene

HER2+ (Human Epidermal Growth Factor Receptor 2) breast cancer cells encode the ERBB2 gene, which translates the HER2 or neu protein. This protein is similar in structure to human epidermal growth factor (HER1) and is a member of the epidermal growth factor family. HER2 is involved in a signal transduction pathway that regulates cell proliferation and apoptosis, thus HER2 positive cells display uncontrolled cell growth. The gene for this growth factor is located on chromosome 17 and is overexpressed in around 30% of breast cancers (Yu et al., 2007). Patients with a mutated ERBB2 gene have been shown to have a worse prognosis than ER+ patients; however monoclonal antibody treatment with Trastuzumab or Pertuzumab can help to improve this (Romond et al., 2005). Trastuzumab and Pertuzumab bind to the HER2 receptors to block signalling pathways that lead to cell proliferation.

1.24.7 : Breast cancer **subtypes that do not** express **the genes for** estrogen receptor **(ER),** progesterone receptor **(PR) and** Her2/neu

Breast cancers that are ER-negative, PR-negative and with show no detectable HER2/neu overexpression are referred to as triple negative (TN) or basal-like (BL). Some basal-like breast cancers are sometimes defined as being slightly different genotypically and phenotypically from triple negative breast cancers, however both types of cancer often come under the same definition and both are very heterogeneous for of gene expression profiling. Basal-like cancer cells are so named because they can display an increase cytokeratin expression such as CK5/6 found in the basal-myoepithelial layer of breast tissue. TN and BL subtypes have been associated with poor prognosis (Banerjee et al., 2006; Hugh et al., 2009) and tend to be more aggressive (Dent et al., 2007) in terms of treatment and survival time after diagnosis when compared with other receptor positive

subtypes.

1.24.8: Expression of tumour markers

Specific DNA mutations or gene expression profiles are identified in the cancer cells this may give an indication of which treatment options are most suited to the patient, either by targeting these changes, or by predicting from the DNA profile which non-targeted therapies are most effective.

1.25: DNA classification systems:

The development of a range of screening assays for the detection of breast cancer coupled with the increased awareness of the disease through media campaigns has allowed a greater number of women to be diagnosed with breast cancer at an early stage. However, regardless of this around 30% of women with small tumours that are negative for lymph node involvement (i.e. no metastasis) suffer a relapse and die from their disease (Wong et al., 1992; **Cardoso et** al., 2007). This means that it is difficult to stratify patients into high and low risk categories and a high proportion of patients receive adjuvant treatment that may not require such toxic treatment (**Cardoso et** al., 2007). It is thus important to decipher which patients need adjuvant therapies.

1.25.1: DNA microarray technology

A difference in the expression of hundreds of genes has been found when comparing breast cancer cells with normal cells using microarray technology. Microarrays allow the expression

levels of large numbers of genes to be measured simultaneously. The significance of differential gene expression is unclear, however it has been shown that clusters of genes co-express, for example ER positive tumours and ER negative tumours (Graham et al., 2010). Multigene assays could allow the refinement of treatment options and give indications of prognosis depending on gene expression signatures. Various types of gene expression profiling by microarray exist such as the American Society of Clinical Oncology (ASCO) approved Oncotype DX and FDA approved MammaPrint.

1.25.1: Oncotype DX

Oncotype DX is a 21 gene assay; 16 breast cancer-related genes and 5 normal 'reference' genes are assessed for ER positive, lymph node negative tumours only. The resulting gene expression profile is given a recurrence score (RS) and a high RS denotes a poor prognosis (Paik et al., 2004).

1.25.2: AmpliChip CYP450

AmpliChip CYP450 from Roche is one of the first FDA approved DNA microarray based tools for the prediction of patient specific phenotypes in relation to drug toxicity and efficacy in treating breast cancer. The technology is based around the cytochrome P450 family, specifically two highly polymorphic genes involved in the metabolism of 25% of all prescription drugs: CYP2D6 and CYP2C19 (Heller et al., 2006). Microarrays detect genotypic differences in these genes and predict 4 possible phenotypes, ranging from poor to hyper rapid, based around 31 polymorphisms and mutations associated with these metabolism associated genes (Heller et al., 2006, de Leon et al., 2006).

1.25.3: MammaPrint

MammaPrint screening is a gene assay that can be used irrespective of ER status in lymph

node negative patients. It was first developed in Amsterdam by Agendia. The expression levels of 70 genes are analysed and an algorithm is calculated allowing the patient sample to be given a score that determines whether the patient is classified as high or low risk for cancer metestasis (Van't Veer et al 2002 and 2006). Therapy can thus be stratified for therapeutics based on their risk of recurrence, administering more aggressive treatment on the high risk patients.

1.26: Breast Cancer Predisposition genes:

High penetrance genes:

1.26.1: BRCA1

Mutations in either of two major susceptibility genes, breast cancer susceptibility gene 1 (BRCA1) and breast cancer susceptibility gene 2 (BRCA2), confer a lifetime risk of breast cancer of between 60 and 85 percent and a lifetime risk of ovarian cancer of between 15 and 40 percent. However, mutations in these genes account for only 2 to 3 percent of all breast cancers.

BRCA1 (breast type 1 susceptibility protein) was first identified in 1994 (Miki et al, 1994). The BRCA 1 gene is located on the long arm of chromosome 17. Mutations in the BRCA genes exhibit high-penetrance, and confer a greater than 10-fold increase in breast cancer risk. BRCA 1 has been shown to be involved in the repair of DSB by homologous recombination by binding directly to DNA, and also via interactions with RAD51. BRCA1 was first shown to be a key player in DNA repair when the association of BRCA1 with RAD51 in nuclear foci was observed (Scully et al, 1997). RAD51 is involved in locating homologous regions and pairing up of DNA strands during the repair process. BRCA 1 has been implicated in the signalling response to DNA damage to ensure faithful DNA repair.

Modulation of BRCA 1 by ATM, ATR and CHK2 has been observed. ATM is recruited and activated by DNA ds breaks. ATM has been shown to phosphorylate BRCA 1 when exposed to gamma radiation (Cortez et al, 1999). ATR, related to ATM, is involved in detecting DNA damage, particularly single stranded regions of DNA indicative of stalled replication forks, an intermediate in DNA repair. ATR has also been shown to associate with and phosphorylate BRCA 1 after cells are exposed to DNA damaging agents (Tibbetts et al, 2000). CHK2, a downstream target of ATM, phosphorylates a serine residue of BRCA 1 and is important in promoting error free homologous recombination and simultaneously inhibiting error prone non-homologous DNA repair mechanisms such as microhomology mediated end-joining (Zhang et al., 2004).

1.26.2: BRCA2

BRCA2 (breast type 2 susceptibility protein BRCA 2 was also discovered in 1994 (Wooster et al, 1994) and is located on the long arm of chromosome 13. BRCA2 has, like BRCA 1, been shown to interact with RAD51 to regulate both the intracellular localisation and DNA-binding ability of this protein. Cells defective in BRCA2 show that BRCA2 is also implicated in DNA repair. In culture mouse embryonic fibroblasts (MEFs) with truncated copies of BRCA2 show defects in cell proliferation and a heightened sensitivity to genotoxic agents such as UV damage (Patel et al, 1998). There is evidence of a range of chromosomal instabilities in cells lacking functional copies of BRCA2 such as DNA breaks (Lee et al, 1999). BRCA2 has also been implicated in successful mitosis; mitotic checkpoint protein inactivation coupled with BRCA2 deficiency causes cells to overcome growth arrest and initiate neoplasms (Lee et al, 1999).

By influencing DNA damage repair, these proteins thus play a key role in maintaining the stability of the human genome. Recently the discovery of PARP inhibitors has shown promise in BRCA deficient cells. As mentioned previously PARP 1 is a DNA repair protein and cells deficient in PARP have increased DSBs. Studies have shown that PARP inhibitors have selective activity for tumours with defects in BRCA1 or BRCA2 (Kaelin, 2005). Cells deficient in BRCA1/2 exhibit dysfunctional DNA double strand break (DSB) repair (e.g. HR or NHEJ). PARP inhibitors are thought to increase cell death in cells with dysfunctional DSB repair via synthetic lethality. Synthetic lethality is when a mutation of either of two genes has no phenotypic effect, but combining the mutations results in cell death (Dobzhansky, 1946; Telli and Ford, 2010). This effect implies that inhibiting one of these genes in a context where the other is defective should be selectively lethal to the tumor cells but not toxic to the normal cells, potentially leading to a large therapeutic (Telli and Ford, 2010).

1.26.3: p53

A key gene that is frequently mutated in breast cancer is the p53 gene. P53 is a tumour suppressor protein encoded by the TP53 gene in humans; it is located on the long arm of chromosome 17. P53 is the most commonly mutated gene in human cancers, mutated in as many as 50% of all human cancers (Hermandez-Boussard et al., 1999; Cheah et al., 2001).

As a tumour suppressor, p53 can activate DNA repair proteins such as ATM; can arrest cell growth; or initiate apoptosis. In a normal cell p53 is present at very low levels due to interactions with its negative regulator mdm2 with which p53 forms a complex. Upon DNA damage or other stresses, various pathways will lead to the dissociation of the p53 and mdm2 complex. DNA damage triggers the activation of protein kinases such as ATM and CHK2,

which phosphorylate p53. Once activated, p53 may induce a cell cycle arrest to allow repair and survival of the cell via the expression of p21. P21 is a downstream target of p53 and acts as an inhibitor of cyclin-dependent kinases that are required for transition in the cell cycle (eg.cdk2). Alternatively p53 may induce apoptosis by triggering the expression of pro-apoptotic proteins such as BAX (Bcl-2 associated X protein).

It has been observed by immunohistochemistry that roughly 40% of breast carcinomas show high levels of stabilised, often mutant, p53 protein in their cells (Barnes et al., 1996). P53 dysfunction can arise via mutagens or can be inherited. The inheritance of just one functional copy of the p53 gene is classified as the rare Li-Fraumeni syndrome. Li-Fraumeni is a cancer pre-disposition syndrome that is associated with an early onset of breast cancer (Fraumeni et al., 1990). Individuals with Li-Fraumeni syndrome possess either a mutant p53 gene or CHK2 gene (CHK2 regulates p53) and this increases susceptibility to cancer as only one allele has to lose functionality to cause dysfunctional cell growth.

1.27: Telomere length in Breast Cancer

A shorter telomere length has been correlated with breast cancer compared with normal tissue in multiple studies. In one example, Heaphy *et al*. focussed on evaluating mean telomere lengths in different cancer subtypes based on tumour characteristics. Using *FISH* at the single cell level it was found that telomere lengths were shorter in more aggressive subtypes, such as luminal B, HER-2-positive and triple-negative tumours, suggesting that tumour telomere length may have utility as a prognostic or risk marker for breast cancer (Heaphy et al., 2011).

1.28: Genomic instability in Breast Cancer

Microarray-based gene-expression profiling, has enabled the enhancement of molecular classification and has also shown how heterogeneous the disease is. Genomic alterations arising from types of genomic instability would explain one mechanism for the extreme variation in breast cancer subtypes and the severity of tumour types. Dysfunction in housekeeping genes early on in breast cancer progression is thought to lead to the accumulation of mutations resulting in large scale genomic instability (Loeb et al., 2001).

Hallmarks of genomic instability visible via *FISH* technology, such as an increase in anaphase bridges, gains and losses of genetic material etc., have been correlated with the progression to malignancy in many types of cancer. Gains and losses on certain chromosome arms in particular are consistent with the presence of breast cancer oncogenes such as those found on 17q (HER2) (Futreal et al.,1992; Bowcock et al., 1993).

Genome instability arising as a consequence of telomere crisis is considered to have a crucial role in the development of most breast carcinomas, in particular during transition from UDH (usual ductal hyperplasia) to DCIS (ductal carcinoma *in situ*) (Chin et al., 2004). Genomic instability and telomere length were observed by Chin *et al.*, using *FISH* in normal ductal epithelium, UDH, DCIS and invasive cancer. Genome instability was assessed phenotypically by confocal microscopy analysis of tissue sections. There were an increase in anaphase bridges detectable in UDH and DCIS. Telomere length analysis showed that telomere length is significantly shorter in invasive cancers than DCIS and shorter in DCIS than UDH. A correlation has thus been shown between genome instability, telomere length and progression to invasive breast cancer.

Comparative Genome hybridisation (CGH) techniques used in breast cancer commonly use

microarray based technologies (Array-CGH) (Pinkel et al., 1998; Pollack et al., 1999) to compare DNA copy number alterations (CNA) between cancer tissue/tumour samples to known genomic sequence positions using oligonucleotide probes. Array-CGH can thus be used to assess single nucleotide polymorphisms (SNPs) and loss of heterozygosity (LOH) as a measure of the extent genomic instability in cancer tissues. CGH techniques have been used as a way to classify breast tumours, for example the distinct genomic profile of BRCA1 tumours, based on losses or gains at chromosomal loci (Wessels et al., 2002; van Beers et al., 2005). Therapeutic targets have been investigated using array-CGH to screen for genes with abnormal DNA copy numbers and to elucide complex patterns of genomic changes accosiated wtih known oncogenes such as ERBB2 and Cyclin D1 (Hosokawa et al., 1998).

1.29: Telomere dynamics in Chronic Lymphocytic Leukeamia (CLL)

The development of a high resolution single molecule approaches enabled the determination of the full spectrum of telomere length in human tissues as well as the detection of rare telomere fusion events (Baird *et al.,* 2004). These techniques were applied to patient samples with CLL (Lin *et al.,* 2010). CLL is the most common form of adult leukaemia (Landis et al., 1998). NRTs are common events in CLL and independent markers of a poor prognosis, along with mutations in p53 and ATM genes on 17p and 11q respectively (Mitelman et al., 1997; Verstovsek et al., 2004). Telomere length at XpYp and 17p was shorter in later Binet's stages of CLL, thus indicating that short telomere length could be indicative of poorest prognosis (Lin et al., 2010; Jones et al., 2012). Interestingly subsets of telomeres in the stage A patient group displayed telomere mean lengths (XpYp and 17p) reminiscent of those found in cells undergoing crisis, and similar to subsets found in stage C. This could indicate early evidence

of patients that will go on to develop a more aggressive form of CLL. Alongside this the shortest, most eroded telomeres here were the ones undergoing the highest frequency of telomere-telomere fusions, a hallmark of telomere dysfunction regardless of stage (Lin et al., 2010). Array-CGH was used to display evidence of genomic instability via CNAs in patients that exhibited telomere dysfunction (Lin et al., 2010). Thus telomere shortening and telomere dysfunction have been documented in human haematological malignancy and correlate with the emergence of increased genomic instability (Lin et al., 2010; Jones et al., 2012).

1.30 Measuring telomere length

1.30.1 Telomere Restriction Fragment (TRF) analysis

Telomere Restriction Fragment (TRF) analysis is the most commonly used assay (Allshire et al., 1989) involving digestion of the DNA with one or two frequently cutting restriction enzymes such as *HinfI* and *RsaI* in human samples (Moyzis et al., 1988). The idea of using frequently cutting enzymes such as these is that digestion of the DNA cleaves the majority of the DNA and leaves the terminal restriction fragments intact to be detected with telomere repeat containing probes. The problem with this method of telomere length detection is that it measures both the telomere repeats and the subtelomere. The hybridisation technology may also limit this method's ability to detect short telomere repeat arrays (Baird, 2005). Furthermore this technique necessitates large amounts of DNA material and is therefore not always suitable for samples with less than 10^5 cells.

1.30.2 Quantitative fluorescence in situ hybridisation (Q-FISH)

Q-FISH technology uses fluorescently labelled (Cy3 or FITC) peptide nucleic acid or

PNA oligonucleotides to quantify $(TTAGGG)_n$ telomere repeat sequences. PNA is a synthetic DNA molecule that contains no charged phosphate groups and binds more readily to DNA. Q-FISH has the ability to study telomere lengths of each individual chromosome (p or q arm) and requires less than 30 cells (Lansdorp et al., 1996). The ability to distinguish chromosome arms led to the finding that specific chromosome ends, for example 17p, appear to be shorter than others (Martens et al., 1998). Q-FISH relies on metaphase chromosomes or interphase nuclei, which limits the cells that can be analysed to those capable of growth (not senescent cells). Also the labelling with fluorescent probes can be inaccurate and unable to detect very short telomeric repeats (Lansdorp et al., 2001; Baird, 2005).

1.30.3 Flow-FISH

An alternative FISH technique to Q-FISH called 'Flow-FISH' was developed in the same laboratory that uses interphase cells and hybridizes the PNA probes in suspension (Rufer et al., 1998; Lansdorp et al., 2006). The use of flow cytometry coupled with hybridisation techniques allows thousands of cells to be analysed in a reasonably short time and has been widely used to study telomere length in haematopoietic system. One drawback of this method is that Flow-FISH detects an average telomere length for each cell analysed not the telomere length of an individual chromosome and because it is hybridisation based it is limit in resolution.

1.30.4 Telomere Q-PCR

Another fluorescence-based assay that uses Q-PCR is 'Telomere Q-PCR (Cawthon, 2002). Telomere Q-PCR measures the ratio of telomere repeat copy number to single copy gene copy number giving a 'T/S ratio' (Cawthon, 2002). This technique was initially compared with TRF analysis for suitability as a telomere length measurement assay, and similar to TRF

only average telomere length can be quantified per sample and individual telomeres cannot be quantified (Cawthon, 2002; Vera and Blasco, 2012). The high throughput of this technique mean it is has been widely used in telomere based epidemiological studies.

1.30.5 Single TElomere Length Analysis (STELA)

In order to overcome drawbacks with fluorescent related techniques, the PCR-based technique Single TElomere Length Analysis was developed. STELA is a high resolution single molecule strategy that amplifies single telomeres within a sample. A linker repeat unit targets the 3′ overhang structure of the telomere terminus and is ligated to the 5′ of the telomere which acts as a unique sequence tag (Figure 1. 8). PCR utilises a tail-specific primer 'teltail' and a chromosome specific upstream primer. Amplification between 'teltail' and the telomere/chromosome specific primer is dependent upon synthesis of the complementary strand of the telorette linker in the first round of PCR. This together with detection of the amplified products by Southern hybridisation means that the length of specific chromosome ends can be determined. STELA was originally developed at the Xp/Yp telomere, but has since been extended to other chromosome ends (Britt Compton et al., 2007).

A key difference between STELA and TRF is that TRF is only capable of determining the mean telomere length of all chromosome ends from a large number of cells, whereas STELA is able to look at individual telomeres; this results in a simplification of the telomere length distributions and an improvement in resolution (Baird et al., 2005). STELA and TRF were compared using human fibroblast strains *in vitro* showing the inability of TRF to detect short telomeres in senescent cell populations (Baird et al., 2003). The key advantage of STELA is its ability to detect short telomeres that have eroded to the point at which they become dysfunctional and capable of fusion (Capper et al., 2007). This has allowed the definition of a

'threshold' length below which telomeres are dysfunctional (Capper et al., 2007; Lin et al., 2010).

Figure 1.8: Representation of STELA:

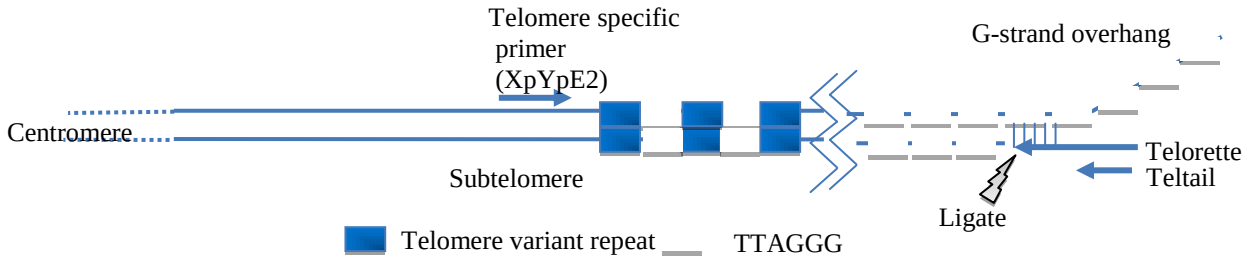


Figure 1.8: STELA PCR: A 'telorette' linker has been designed that is complementary to 7bp overhang; the Teltail primer homologous to overhang allows amplification between telomere specific primer and teltail Oligonucleotide primers and linkers are shown by blue arrows. Adapted from (Baird et al., 2003)).

Project Aims:

This project involved technology development *in vitro,* together with translational work, aiming to use novel single molecule strategies to show that telomere dysfunction occurs inbreast cancer.

The first aim of this project was to further develop single-molecule telomere fusion technology; it was hoped that this would increase the sensitivity of the assay and allow the detection of more complex mutational structures than could be detected previously.

Secondly telomere length analysis was carried out using high resolution single molecule PCR strategies (STELA) in a panel of DNA samples derived from invasive ductal carcinoma. This aimed to look for evidence of extreme telomere shortening to below the length threshold that had been previously determined as critical for the onset of telomere fusion.

A third aim was compare the telomere length data with clinical data received for all samples denoting receptor status, morbidity among other characteristics, with a view to determine the potential utility of telomere length as a potential prognostic indicator in breast cancer.

The final aim was to apply single-molecule telomere fusion assays to clinical samples to examine if telomeres within breast cancers are capable of undergoing telomere-telomere fusions that have the potential to drive genomic instability and the progression to malignancy.

Chapter 2: Materials and methods:

2.1 Chemicals and Reagents

Chemicals were obtained from Fischer Scientific (Loughborough, UK) and molecular biology reagents obtained from GObco BRL/Invitrogen Ltd (Paisley, UK), Stratagene (La Jolla, California), ABgene (Surrey, UK), Promega (Southhampton, UK), New England Biolabs (UK) Ltd (Hertfordshire, UK), Roche products Ltd (Hertfordshire, UK) Radiochemical reagents were obtained from Amersham biosciences/GE Healthcare Ltd (Buckinghamshire, UK) and PerkinElmer Life and Analytical Science (Milano, Italy).

2.2 Plastic laboratory equipment

Plastic labware and glassware used for experiments was obtained from Gilson, Becton Dickinson labware, Eppendorf and Thermo Scientific.

2.3 Equipment/Machinery

Equipment for use in experiments was obtained from: Bio-RAD, EPS, Finn, Gilson, MSE, Hybaid, Amersham, Thermo scientific and Qiagen. Specialised equipment consisted of: PCR thermocycler (Bio-RAD), Transilluminator (EPS), Centrifuge (MSE), Hybridisation ovens (Hybaid), Gel electrophoresis tanks (Amersham), Fluorometer (Amersham), DyeEX spin kits (Qiagen) and Pipettes (Finn and Gilson)

2.4 Oligonucleotides used:

Oligonucleotides were designed using sequence data for telomere ends compiled by H. Riethman at the Wistar Institute, (URL: // http://www.wistar.upenn.edu/Riethman/). The subsequent oligonucleotides were then synthesised by MWG-biotech AG (Ebersberg, Germany).

List of Oligonucleotide used for this project:

2.4.1 STELA Oligonucleotides

Name	Sequence
Teltail	5’-TGCTCCGTGCATCTGGCATC-3’
Telorette2	5’-TGCTCCGTGCATCTGGCATCTAACCCT-3’

2.4.2 XpYp telomere primers

XpYpB2	5’-TCTGAAAGTGGACC(A/T)ATCAG-3’
XpYpC	5’-CAGGGACCGGGACAAATAGAC-3’
XpYpE2	5’-TTGTCTCAGGGTCCTAGTG-3’
XpYpG	5’-AATTCCAGACACACTAGGACCCTGA-3’
XpYpM	5’-ACCAGGTTTTCCAGTGTGTT-3’
XpYpO	5’-CCTGTAACGCTGTTAGGTAC-3’
XpYpP	5’- ACCAGGGGCTGATGTAACG-3’
XpYpQ	5’-CCATGAGACACAAAGGACTC-3’
XpYp433AT	5’-GGTTATCGACCAGGTGCTCT-3’
XpYp433GC	5’-GGTTATCGACCAGGTGCTCC-3’

2.4.3 17p telomere primers

17p6	5’-GGCTGAACTATAGCCTCTGC-3’
17pseqrev1	5’-GAATCCACGGATTGCTTTGTGTAC-3’

2.4.4 21q Telomere primers

21q1	5’-CTTGGTGTCGAGAGAGGTAG-3’
21qseq1rev	5'-AGCTAGCTATCTACTCTAACAGAGC-3’

2.4.5 16p telomere primers

16pseq1rev	5'-GCTGGGTGAGCTTAGAGAGGAAAGC-3

2.4.6 9p telomere primers

9p1	5’-TGCGTTCTCGTCAGCACAGACCC-3’
9p2	5’-CACATTCCTCATGTGCTTACG-3’

2.4.7 ALU primers

ALU FORWARD	5’-TGGCTCACGCCTGTAATC-3’
ALU REVERSE	5’-GAGCGAGACTCCGTCTCA-3’

2.4.8 LINE-1 primers

L1	5’-GAACAGCTCCGGTCTACAGCTC-3’
L2	5’-GCGTGAGCGACGCAGAAGAC-3’
L3	5’-CAGGCCTCCTTGAGCTGTGGTG-3’
L4	5’-GTCTGCAGAGGTTACTGCTGTC-3’

2.4.9 Adaptor-ligation mediated PCR

Long strand of adapter 1	5’-GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGC-3’
Short strand of adapter *	5’-Phosphate-AATTGCAGCCCG-amino C7-3’
Adapter primer 1 (AP1)°	5’-GTAATACGACTCACTATAGGGC-3’
Adapter primer 2 (AP2) Δ	5’-TGGTCGACGGCCCGGGCTGC-3’

*HPLC purified; 5’ phosphorylated and 3’ C7 amino modification. °Primer for first PCR.

ΔPrimer for second nested PCR

A1	5’-Phosphate-AATTCATGATGCCAGATGCACGGAGCA-3’ (HPLC)
A2	5’-AATTCATGATGCCAGATGCACGGAGCA
A3S	5’-AATTCATGATGCCAG-3’
A4S	5’- Phosphate-AATTCATGATGCCAG-3’ (HPLC)

ECOR1 LINKER 1	5’-TGCTCCGTGCATCTGGCATCATG-3’
ECOR1 PRIMER	5’-CATCTGGCATCATGAATTCCTG-3’

2.4.10 Splinkerette primers

Splink1	5’-CGAAGAGTAACCGTTGCTAGGAGAGACC-3’
Splink2	5’-GTGGCTGAATGAGACTGGTGTCGAC-3’
SP1	5’- GACACTAGTGGGATCAGCTC-3’
SP2	5’-GTGGTCTAGGGTGATCACAG-3’

2.4.11 Adaptor sequences for splinkerette ligation

Long-strand adaptor	5’-CGAAGAGTAACCGTTGCTAGGAGAGACC-GTGGCTGAATGAGACTGGTGTCGACACTAGTGG-3’
Short-strand adaptor	5’-GATCCCACTAGTGTCGACACCAGTCTCTAA-TTTTTTTTTTCAAAAAAA-3’

2.5 Samples

DNA samples extracted from breast cancer tissue Invasive Ductal Carcinoma (IDC) were obtained from the Wales Cancer Bank. Microtubes of DNA were received in single tube, in duplicate and in triplicate. The samples that were received in multiple tubes were pooled into single microtubes and labelled from 1-30 for the first set of samples and from 1-100 for the second set. List of patient samples (in a table):

Sample I.D.	
RR6BL0000011FT	1
RR6BL0000067FT	2
RR6BL0000045FT	3
RR6BL0000072FT	4
RVCC40000632FT	5
RR6BL0000002FT	6
RR6BL0000004FT	7
RR6BL0000013FT	8
RR6BL0000001FT	9
RR6BL0000039FT	10
RVCC40000573FT	11
RVCC40000166FT	12
RVCC40000649FT	13
RVCC40000589FT	14
RVCC40000539FT	15
RVCC40000033FT	16
RR6BL0000008FT	17
RVCC40000030FT	18
RVCC40000755FT	19
RVCC40000165FT	20
RVCC40000042FT	21
RVCC40000025FT	22
RVCC40000009FT	23
RVCC40000155FT	24
RVCC40000051FT	25
RVCC40000583FT	26
RVCC40000144FT	27
RVCC40000647FT	28
RVCC40000758FT	29
RVCC40000698FT	30

Sample	I.D.	Number
RR6BL	20	1
RR6BL	33	2
RR6BL	83	3
RR6BL	104	4
RR6BL	111	5
RR6BL	117	6
RR6BL	120	7
RR6BL	131	8
RR6BL	134	9
RR6BL	170	10
RR6BL	171	11
RR6BL	184	12
RR6BL	186	13
RR6BL	198	14
RR6BL	199	15
RR6BL	202	16
RR6BL	204	17
RR6BL	205	18
RR6BL	213	19
RR6BL	214	20
RR6BL	225	21
RR6BL	230	22
RR6BL	232	23
RR6BL	235	24
RR6BL	245	25
RR6BL	252	26
RR6BL	274	27
RR6BL	350	28
RR6BL	386	29
RVCC	348	30
RVCC	965	31
RVCC	1036	32
RVCC	1055	33
RVCC	1469	34

Sample	I.D.	Number
RVCC	1	35
RVCC	16	36
RVCC	21	37
RVCC	22	38
RVCC	26	39
RVCC	29	40
RVCC	31	41
RVCC	70	42
RVCC	72	43
RVCC	89	44
RVCC	129	45
RVCC	135	46
RVCC	665	47
RVCC	907	48
RVCC	642	49
RVCC	559	50

I.D.	Number
VCC 643	51
VCC 090	52
VCC 020	53
VCC 631	54
6BL 259	55
6BL 623	56
6BL 064	57
6BL 141	58
VCC 936	59
VCC 067	60
VCC 154	61
VCC 076	62
VCC 786	63
VCC 1236	64
6BL 409	65
6BL 1111	66
VCC 1060	67
VCC 201	68
6BL 571	69
6BL 130	70
6BL 528	71

VCC 540	72
6BL 535	73
6BL 557	74
6BL 070	75

VCC 597	76
VCC 696	77
VCC 134	78
6BL 005	79
VCC 035	80
VCC 737	81
VCC 071	82
VCC 555	83
VCC 591	84
VCC 795	85
VCC 638	86
VCC 039	87
VCC 633	88
6BL 071	89
VCC 015	90
VCC 687	91
VCC 011	92
6BL 019	93
VCC 686	94
VCC 753	95
VCC 124	96
VCC 622	97
VCC 063	98

VCC 752	99
VCC 776	100

2.6 DNA extraction

DNA was extracted from HEK-293, MCF7 and MRC5 E6/E7 cells using standard proteinase K, RNase A, and phenol/chloroform. Cells were lysed by the addition of 300 µl of lysis buffer (0.5% SDS, 100 mM NaCl and 10 mM Tris-HCL). To this suspension, 60 µg of Proteinase K and 30 µg of RNase was added. Cells were then incubated on a hotblock at 45°C overnight. Cells were spun quickly in a benchtop centrifuge and then 300 µl of phenol/chloroform was added, and the cell suspension was rotated on a tube rotator for 30 minutes. Phase separation was achieved by centrifugation at 13,000 rpm for 5 minutes. Aqueous and interphase phases were pipetted into a fresh eppendorf containing 300 µl of phenol/chloroform. This mix was rotated again for 20 minutes and centrifuged (13 K rpm) for 5 minutes. At this stage, only the aqueous phase was transferred to a fresh eppendorf. 30 µl of 3M NaOAc (pH 5.3) was added to the aqueous phase and mixed (gently). Ice cold 100 % Ethanol (900 µl) was added to precipitate the DNA and the mix was left for one hour

(minimum) at -20C. The DNA suspension was then centrifuged for 1 minute (13K rpm), and this resulted in the formation of a pellet. The ethanol (100%) was removed and 70% ice cold ethanol was added to wash the pellet and the tube was left on ice for 10 minutes.

2.7 Quantification of DNA

DNA samples were quantified in triplicate by Hoechst 33258 fluorometry (Bio-Rad) as detailed previously (Baird et al. 2003). A selection of breast cancer DNA samples provided by the Wales Cancer bank arrived with diluted DNA concentrations (less than 20ng/ul) and showed variable to no reading using the fluorometer; these samples were subsequently quantified in triplicate using Nanodrop-1000 technology supplied by the Central Biotechnology Services (CBS) Core Facility in the School of Medicine, Cardiff University (www.cardiff.ac.uk/medic/cbs).

2.8 STELA

The ligated DNA was diluted to 250 pg μl−1 in 10Mm Tris-HCL pH 8.0 and then multiple PCR reactions were carried out (typically 6 reactions per sample) for each test DNA in volumes of 10 μl containing 100–250 pg of ligated DNA, 0.5 μM telomere-adjacent and teltail primers, 1.2 mM NTPs, 75 mM Tris-HCl (pH 8.8), 20 mM (NH4)2SO4, 0.01% Tween-20, 1.5 mM MgCl2 and 1 U of a 25:1 mixture of *Taq* (ABGene) and *Pwo* polymerase (Roche Molecular Biochemicals).

The reactions were cycled with an MJ PTC-225 thermocycler (MJ research) under the following conditions: 25 cycles of 94 °C for 15 s, 65 °C (XpYpE2) or 66.5 °C (XpYp-427G/415C and XpYp-427A/415T allele-specific primers) for 30 s, and 68 °C for 10 min.

2.9 TVR PCR

The following reagents were used per reaction PCR reaction tube: 100ng DNA, 1.2 mM NTPs, 75 mM Tris-HCl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 1.5 mM $MgCl_2$, 0.5

U of Taq polymerase, 1.0µM telomere adjacent primer XpYpE2 and 1.0µM TVR primers Tel-X, Tel-W or Tel-Y. The reactions were cycled under the following conditions: 33 cycles of 15 sec at 94°C, 30 sec at 65°C, and 3 mins at 70°C.

2.10 Allele specific STELA

The same protocol was employed as used for 'STELA' PCR, except allele specific primers XpYp433AT and XpYp433GC were used individually in the place of the telomere adjacent primer.

2.11 Fusion PCR

DNA extractions were diluted to a final concentration of 50 ng/µL in 10 mM Tris-HCl pH 8. Multiple PCRs were then carried out (typically 9 reactions per primer combination) using 50ng/ul DNA, 0.5 µM telomere-adjacent primers, 1.2 mM NTPs, 75 mM Tris-HCl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 1.5 mM $MgCl_2$, and 0.5 U of a 10:1 mixture of Taq and Pwo polymerase. The reactions were cycled under the following conditions: 25 cycles of 15 sec at 94°C, 30 sec at 59°C, and 8 min at 68°C.

2.12 Standard PCR

The following reagents were used per PCR reaction tube: 100ng/ µl DNA, 0.5 µM forward primer, 0.5 µM reverse primer,1.2 mM NTPs, 75 mM Tris-HCl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 1.5 mM $MgCl_2$, and 0.5 U of Taq polymerase. The reactions were cycled under the following conditions: 33 cycles of 15 sec at 94°C, 30 sec at 60°C, and 1 min at 68°C.

2.13 DNA digestion

2µg of genomic HEK-293 DNA was digested in a reaction mix containing 40 U *Sau*3A1 enzyme; 1/10 volume of 10x NEB buffer 1; 0.33µg/µl Acetylated BSA and 23µl double

distilled H_2O with a final reaction volume of 30µl. The reactions were carried out in triplicate and incubated at 37°C for 12 hours. This was immediately followed by an enzyme denaturation step of 65°C for 20mins. The digested DNA was then stored at 4°C or at -20°C indefinitely.

2.14 Preparation of Oligonucleotides for *Eco*RI adaptor ligation mediated PCR

Adaptors were designed using XpYp sequence data generated by H. Riethman at the Wistar Institute, (URL: // http://www.wistar.upenn.edu/Riethman/). A linker was designed to target a specific *Eco*RI site at the XpYp telomere located 146bp from the telomere end. To test for ligation efficiency a variety of different 'adaptors' were designed that were different lengths (10 or 20bp) and either phosphorylated or unphosphorylated at the 5' terminus. The adaptors were designed to be specific to the adaptor sequence and were annealed in various combinations (see Figure 2.14 and Table 2.14) using a protocol designed by Artandi *et.al* 2005. 20µl of each oligonucleotide (5µM concentration) was added to 1mM Tris-HCl pH 8.3. The mix was then heated to 96°C for 2 minutes and cooled gradually to room temperature at a rate of 5°C/minute.

Figure 2.14

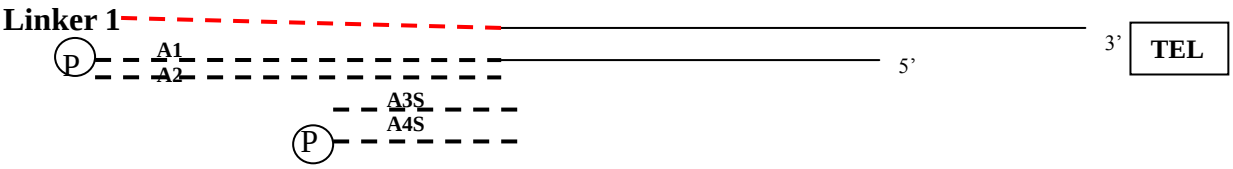


Figure 2.14: Diagrammatic representation of the orientation of linkers and adaptors that have been designed at the XpYp telomeric end.

Table 2.14

Combinations:

Linker 1 (L1	Adaptor 1 (A1)
Plus	Adaptor 2 (A2)
	Adaptor 3S (A3S)
	Adaptor 4S (A4S)

Table 2.14: This table shows the different sets of adaptors to be used in combination with the 'linker 1' oligonucleotide that is specific to an Eco*RI restriction site at XpYp.*

2.15 Ligation of *Eco*RI linkers

Typically 4-6 reactions were performed for each digested of DNA and then pooled to form a stock of ligated DNA. 300ng of *Eco*R1-digested genomic DNA was added to a reaction mix of 25 pmol Adaptor mix (see above); 20U T4 DNA ligase; 1X T4 DNA ligase buffer and 15 µl double distilled H_2O. This reaction was incubated at 16°C for 30 minutes. The T4 DNA ligase enzyme was inactivated by heating each reaction to 65°C for 10minutes. Ligated DNA was stored at -20°C.

2.16 Adaptor ligation mediated PCR (ECORI)

For each PCR reaction tube a mix of 1 µl digested and ligated DNA, 0.5 µM forward primer,

0.5 µM reverse primer,1.2 mM NTPs, 75 mM Tris-HCl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 1.5 mM $MgCl_2$, and 0.5 U of Taq polymerase were used. The reactions were cycled under the following conditions: 30 cycles of 20 sec at 94°C, 30 sec at 60°C, and 5 min at 68°C.

2.17 Preparation of Splinkerette adaptor mix

The reagents were set up as detailed by Uren, A.G. *et al.* A 25µM concentration of each adaptor oligonucleotide was achieved by adding 50µl of 50µM 'long-strand adaptor' to 50µl of 50µM 'short-strand adaptor' in a 1.5ml eppendorf. The mix was stored at -20°C.

2.18 Ligation of Splinkerette adaptor to *Sau*3A1digested DNA

Typically 4-6 reactions were performed for each digested of DNA and then pooled to form a stock of ligated DNA. 300ng of *Sau*3A1-digested genomic DNA was added to a reaction mix of 25 pmol Adaptor mix (see above); 20U T4 DNA ligase; 1X T4 DNA ligase buffer and 29.5 µl double distilled H_2O. This reaction was incubated at 4°C for 12 hours. The T4 DNA ligase enzyme was inactivated by heating each reaction to 65°C for 20mins. Ligated DNA was stored at -20°C.

2.19 Splinkerette PCR

The following reagents were used per reaction PCR reaction tube: 40ng/ µl DNA, 0.5 µM XpYp specific primer, 0.5 µM splinkerette primer,1.2 mM NTPs, 75 mM Tris-HCl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 1.5 mM $MgCl_2$, and 0.5 U of Taq polymerase. The reactions were cycled under the following conditions: 25 cycles of 15 sec at 94°C, 30 sec at 60°C, and 5 min at 72°C.

2.20 Gel electrophoresis for STELA and fusion products

The DNA fragments were resolved by 0.5% Tris-acetate-EDTA agarose gel electrophoresis. PCR products contained 1x Ficoll based loading buffer. For STELA one 40 well comb was used to load each sample and the gel was run along a 40cm gel at 120V for 16hours. For fusion analysis, two to three 40 well combs were used to load the gel and the DNA was run through the same sized 40cm gel at 50V for 16 hours or 140V for 3-4 hours for a good resolution.

2.21 For other PCR products

The DNA fragments were resolved by 0.7% or 1% (for smaller fragments) Tris-acetate-EDTA Agarose gel electrophoresis. Gels were run on a smaller 10-20cm gel at 90V for 1-2 hours.

2.22 Visualisation of PCR products on Agarose gels

Ethidium bromide stained bands were visualised using a UV-transilluminator.

2.23 Southern blotting

Resolved DNA fragments within agarose gels from STELA and fusion PCR were first washed on a shaker in depurination buffer (0.25M HCl) for two 6 minute washes. The gel was then washed in denaturation buffer consisting of 1.5M NaCl/0.5M NaOH for 15 minutes. A positively charged membrane (Hybond-XL, Amersham) was used for transfer of DNA fragments by alkaline Southern blotting in the same denaturation buffer that was used for the final wash.

2.24 Probe labelling and hybridisation

25ng of probe DNA and ladder (equal ratios of 1kb and 2.5kb) in Tris-HCL and 1mM EDTA (TE) buffer was labelled either in a rediprime II reaction or GE ready-to-go bead labelling system (Amersham Biosciences and GE healthcare). Both kits generate radiolabelled probes using random hexaprime labelling with the α-33P to incorporate $^{\alpha}$dCTP into synthesised DNA.

Southern blots were prepared for hybridisation by incubation at 60°C with 15ml church buffer (0.5M sodium phosphate buffer pH7.2) consisting of 1M disodium dihydrogen phosphate and 1M disodium hydrogen phosphate, 1%BSA, 1mM EDTA and 7% SDS, for 30 minutes.

25µl of radiolabelled probe was subsequently added to the hybridisation bottle and left overnight at 60°C.

2.25 Removing unbound probes

A wash buffer was prepared using a mix of 0.1 x sodium chloride sodium citrate/0.1% Sodium Dodecyl Sulphate in double distilled H_2O. Hybridised southern blots were washed several times at 60°C and dried in filter paper at the same temperature.

2.26 Visualisation of Radiolabelled blots

Radiolabelled Southern blots were placed in a cassette with a molecular dynamic phosphoimager screen (Amersham) for 1 hour and subsequently the phosphoimager screen was scanned using the Typhoon 9410 biomolecular imager (GE Healthcare). The screen was then blanked using a 15minute exposure to visible light and the placed back in the cassette for a 24 hour minimum exposure and the blot was scanned again using the Typhoon. STELA blots and fusion-PCR blots were stripped using boiling 0.1% SDS prior to reprobing with a second probe.

2.27 Sequencing

2.27.1 Preparation

PCR products to be sequenced were seeded with dH2O in a 1:20 dilution and inputted into a reamplification PCR which is the standard PCR protocol with 3x all reagents and 33 cycles in the thermocycler. 30ul of the resulting product was ran on a 0.7% agarose gel and purified from the agarose gel using the QIAquick gel extraction kit from Qiagen.

2.27.2 Sequencing reaction

4.4ul of purified DNA suspended in Qiagen's elution buffer (buffer EB) was added to 0.16uM sequencing primer and 4ul of BigDye Terminator Cycle sequencing mix (version 3.1) in the following sequencing reaction: 96oC for 30 seconds, 50oC for 15 seconds and 60oC for 4 minutes.

Products from this thermocyclic reaction were purified using Qiagen's Dye Ex 2.0 spin kit and resulting DNA was run on a sequencing gel by the Central Biotechnology Service (CBS), Cardiff University Heath Park Campus, UK.

2.72.3 Analysis of sequences

Resulting sequencing data was analysed using both sequence scanner (Version 1.0-Applied biosystems) and FinchTV (Software from Geospiza). Sequences that had been edited using this software were subsequently analysed using BLASTN software on the NCBI website (URL:http://NCBI.nih.nlm.gov.uk) to confirm sequence identity against sequence databases.

2.28 Statistics and analysis of gels

Gels that had been scanned using the Typhoon 9410 were subsequently analysed using Molecular dynamics ImageQuant 5.0 (GE). Telomere molecular weights on each STELA Southern blot were calculated using Phoretix 1D software from Nonlinear dynamics. Calculated molecular weights were then exported to Microsoft Excel (2007 version) whereby

the distance between the primer site (primer used in STELA PCR) and the beginning of the telomere repeat array could be subtracted to reveal more accurate telomere lengths.

Mean telomere lengths at XpYp, 17p and 2p for each patient sample, standard deviation, standard error and lower 25th percentile values were calculated in Excel.

General statistics used included a One Way ANOVA were used for ALL comparisons where there are more than 2 groups for example to compare means of telomere length data across three telomere loci for all samples, a Student's t-test to compare telomere means and Bartlett's test for equal variances.

Kaplan Meier survival curves were created using Graphpad Prism^{tm}. Mantel-cox and gehan-Breslow Wilcoxon statistics for survival were used in Kaplan Meier analysis to assess differences in the survival curves generated.

Chapter 3: Technology development: Developing the Telomere Fusion Assay

3.1 Abstract

Single-molecule PCR-based telomere fusion assays have been developed that allow the detection of fusion between a subset of chromosome ends. However, the ability of these assays to detect more diverse events is limited. The aim of the work described in this chapter was to develop fusion assays to screen for fusion events occurring between telomeric and non-telomeric loci, thus increasing the scope of the assay and allowing the detection and characterisation of a wider mutational spectrum.

Three strategies were tested, firstly, repeat elements, such as LINEs and SINEs, have been shown to be located near to common fragile sites and were used as a starting point for a non-telomeric locus that could be used to detect fusions. A combination of Alu and LINE-1 element forward and reverse primers were tested together with the previously characterised telomere fusion assay primers. This technique did not result in an increase in the number of detectable telomere fusion events.

Secondly, two different PCR based genome walking techniques, previously described for the isolation of unknown DNA sequences, were tested as an approach to identify telomere-fusion junctions. Unfortunately, neither inverse PCR, nor adaptor-ligation mediated PCR, yielded fragments that were consistent with telomere fusion events.

The final approach was to test a high throughput ‘splinkerette’ PCR protocol used to screen for transposon inserts within the genome of mice. Optimisation of this technique for use with the human XpYp telomere-adjacent DNA has allowed the amplification of bands that were consistent with single-molecule telomere fusion events. Re-amplification of these bands allowed sequence data to be generated that shows fusion of the XpYp telomere with a region of the 9p chromosome. These indicate that with further development this technique may yield a more diverse range of telomere fusion events.

3.2 Introduction

3.2.1 Telomere fusion assay

A single-molecule PCR-based telomere fusion assay has previously been developed for two chromosome ends, XpYp and 17p. The sub-telomeric sequence of these telomeres is sufficiently unique to allow the design of telomere specific primers for the use of this assay (Baird et al., 1995; Britt-Compton et al., 2006; Capper et al., 2007). Since the development of this assay, primers have been designed to encompass two subtelomeric sequence families: 21q and 16p. This allowed the detection of fusions between the 21q groups, including: 21q, 1q, 2q, 5q, 6q, 6p, 8p, 10q, 13q, 17q, 19p, 19q, 22q and the 2q13 interstital locus; also the 16p group: 16p, 1p, 9p, 12p, 15q, the 2q14 interstitial locus and XqYq telomeres. However, the analysis of fusion between telomeres alone is limiting; increasing the scope of the assay to include more fusion structures will improve the sensitivity and allow the detection of rare fusion events; which will be important in examining telomere fusion in clinical samples where fusion is likely to be rare. It will also allow the detection of a broader range of fusion structures and the identification of specific chromosome ends or 'hotspots' in the human genome that are subjected to fusion.

Three different approaches were taken to improve these telomere-fusion assays.

1. Utilising interspersed repetitive elements
2. Adaptor mediated PCR approaches
3. Splinkerette PCR

3.3 Telomere fusions and fragile sites

Telomere fusion events involving regions close to known fragile sites have been identified, for example fusions involving telomeres and the non-telomeric loci, 2q14, 8q24.3, 7p21.1 and Xp22.1, all of which have been documented as being in the same cytogenetic location at fragile sites (Letsolo et al., 2010; Debacker, and Kooy, 2007). The involvement of fragile

sites has been documented in the process of tumour development via large-scale chromosomal abnormalities, commonly deletions and translocations (Schwartz et al., 1998; Arlt et al., 2006; Glover, 2006; Debacker and Kooy, 2007). And a number of fragile sites are thus associated with human disease and cancer (Debacker and Kooy, 2007), for example at FRA3B is the gene FHIT (fragile histadine triad protein), and encodes an enzyme that acts as a HER2 tumour suppressor (Glover, 2006; Bianchi et al., 2007). Large scale genomic rearrangements could potentially be generated by the fusion of dysfunctional telomeres with a fragile site, facilitating the progression to malignancy.

The sequence of events and the underlying mechanisms that lead to the fusion of telomeres with fragile sites is not clear, i.e. whether there is an initial fusion event between one telomeric end and a fragile site sequesters further fusions with other telomeres or whether telomere fusion generate a fragile site through the creation of inverted repeats when telomeres fuse with sister chromatids (Letsolo et al., 2010). Thus an assay to screen for fusions with fragile sites could help elucidate these mechanisms and potentially understand how telomere fusion could contribute to the mutational events involved in the progression of cancer.

LINES and SINES have been identified in close proximity to common fragile sites such as FRA3B and FRA14A (Arlt et al., 2006; Ragland and Glover, 2009). Therefore one aspect of this project will be to screen for fusions occurring in proximity to loci containing LINES and SINES.

3.4 Long Interspersed Nuclear Elements (LINES) and their role in genomic instability

Retrotransposons are a type of interspersed repeat element that occur in abundance of over 40% in the human genome (Lander et al., 2001). These DNA repeat elements have been postulated to confer genetic variation within the human genome by the ability of some active forms to mobilise via recombinational mechanisms (Pace and Feschotte, 2007). LINEs are a type of retrotransposon element commonly found in eukaryotes encompassing about 17-21%

of the human genome (Lander et al., 2001). Line-2 (L2) elements are thought to be an ancestral and inactive nuclear element, whereas Line-1 (L1) elements are active and can retrotranspose (Smith et al., 1995; Deininger and Batzer, 2002). L1 is a 6-7 kb element (Singer, 1982) and functions as an autonomous retrotransposon, encoding its own reverse transcriptase enzyme facilitating its mobility and proliferation [Reviewed by Mills et al., 2007]. The presence of reverse transcriptase at ORF2 found on L1 elements has been found to amplify other retrotransposons, cause unequal homologous recombination, deletions and insertions within the genome (Dewannieux et al., 2003; Mills et al., 2007). These alterations not only increase the genetic variation of humans but also could implicate these retrotransposons in human disease, such as Duchenne's muscular dystrophy whereby insertions of LINE elements have been shown to affect the protein dystrophin (Narita et al., 1993). LINE insertions have also been shown to play a role in sporadic colon cancer tumour formation via insertions that disrupt the APC gene (Miki et al., 1992). Tumour specific rearrangement of a *myc* locus was found to contain a sequence sharing homology to an L1 (LINE-1) sequence giving evidence for L1 functioning as a mobile element to give rise to a mutation in breast tumour tissue (Morse et al, 1988).

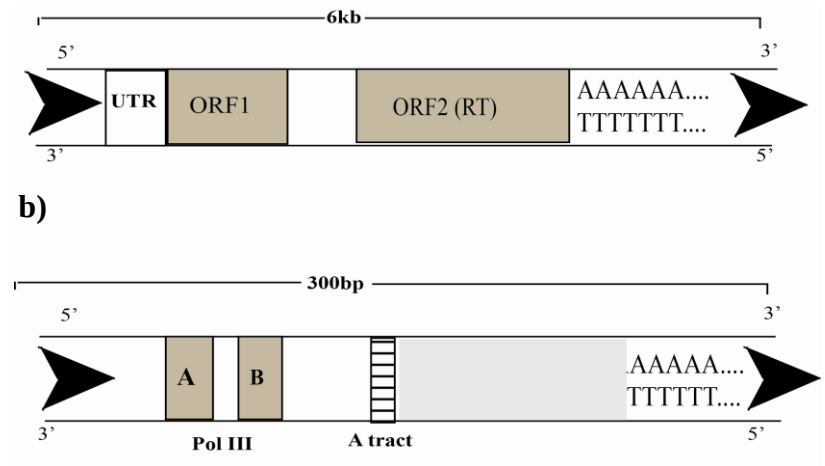
3.5 Short Interspersed Nuclear Elements (SINES) and genome instability

SINES are also abundant in the genome, encompassing around 13% (Lander et al., 2001). Unlike LINEs, SINEs do not encode their own enzymatic machinery for retrotransposition, but instead have been hypothesised to utilise LINE transposase for their own mobilisation, although mechanisms are unclear (Jurka, 1997; Ostertag et al., 2003; Callinan and Batzer, 2006). They tend to be much smaller than LINE elements, spanning up to 500bp (Stansfield and King, 1997). *Alu* elements (~300bp) are a the most common type of SINE found in humans and are so named because of the presence of *Alu*I restriction sites (Houck et al., 1979). *Alu* elements have also been observed to undergo homologous recombinations,

unequal exchange, and cause insertions in the genome Oldridge et al., 1999; Deininger and Batzer, 1999). Events such as this have been linked to genome instability and thus numerous types of human disease (Deininger and Batzer, 1999). *Alu* recombination events have been shown in the All-1 gene, which is mutated in a large number of acute lymphocytic leukaemia cases (Schichman et al., 1994). Breast cancer susceptibility genes have large numbers of *Alu* element sequences (about 42% at BRCA1 and 20% at BRCA2), and rearrangements involving *Alu* sequences have been shown to disrupt the function of both BRCA1 and BRCA2 genes (Miki et al., 1996; Puget et al., 1999).

Repeat elements have thus been demonstrated to act as hotspots for genome instability.

Figure 1: The basic structure of a LINE and SINE



OBJ

Figure 1: a) Shows the basic structure of a LINE-1 element consisting of two ORFs, the second of which encodes a reverse transcriptase for mobilisation **b)** Shows the basic structure of a SINE: The different shades of grey represent two separate 'arms', divided by an A-rich tract. The first arm contains two RNA polymerase III promoter boxes. (Figure adapted from Callinan and Batzer, 2006)

3.5.1 Project aims

The current telomere fusion assay allows the detection of single telomere–telomere fusion molecules between 13 chromosome ends, however there could be 1000s of possible fusions occurring within the genome therefore assay is limited. This project aims to increase the scope of the current fusion assay to detect and characterise a broader mutational spectrum. To do this, interspersed repetitive elements will be used to screen for fusions occurring close to these regions and genome walking technologies will be applied: Adapter mediated ligation using EcoRI restriction sites; Splinkerette PCR using Sau3ai sites.

3.6 Developing a SINE fusion assay: *ALU*-PCR

It was hypothesised that an adapted *ALU* PCR could be used to detect fusions between telomeres and Alu elements by utilising Alu in the existent telomere fusion assay. Primers designed against a human *Alu* sequence were both forward and reverse in orientation, these *Alu* primers were thus used in conjunction with telomere specific primers designed previously. To develop this assay DNA was extracted from HEK-293 cells; these cells exhibit high frequencies of telomere fusion (Capper et al., 2007) and thus provide a good source material to develop new fusion assays.

An annealing temperature gradient PCR was used to optimise the *Alu* primers for use with a combination of primers to be tested in the existing telomere fusion assay conditions using HEK-293 DNA. The PCR reaction was cycled with a temperature gradient ranging between 55°C and 72°C. Table 2 highlights all combinations of primers used in this assay. The clearest bands were visible under ultraviolet light for annealing temperature conditions of between 60°C and 65°C; therefore 60°C was selected for further experiments (data not shown).

3.7 Results: *Alu*-telomere fusions

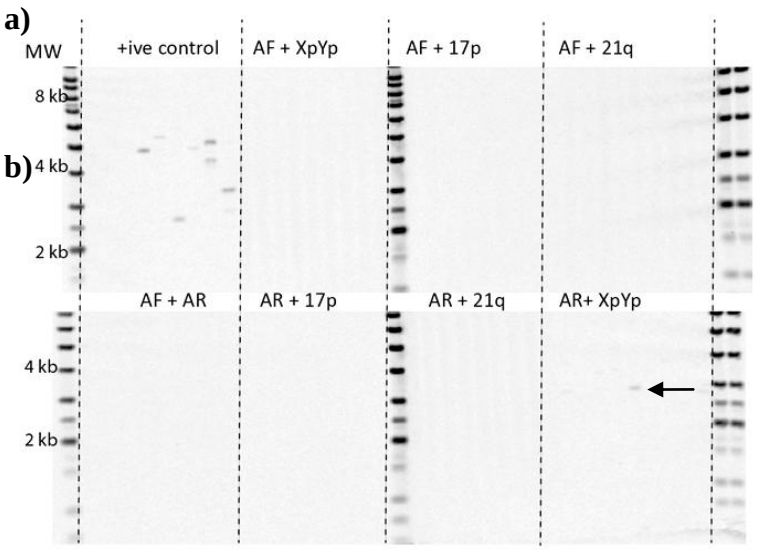
Putative fusion products resulting from PCR amplification using the *Alu* reverse primer and the XpYp specific telomere fusion primer yielded two bands detected by Southern blot with the XpYp o-g probe. (Figure 2). These bands indicate possible fusions that could be occurring between the XpYp telomere and the terminus of an *Alu* element. Other primer combinations yielded minimal results, and the results were not reproducible. No other primer combinations produced bands that were consistent with fusion events when comparing these to fusions visible on a southern blot for the positive HEK-293 control. Attempts to characterise these fragments by PCR re-amplification was unsuccessful thus it was inconclusive whether or not

a fusion between *Alu* elements and telomeres are being detected here due to the inability to sequence the bands.

3.8 Conclusion: Success of the *Alu*-PCR

This technique is limited in the detection of fusion events between telomeres and *Alu* elements in the human genome. This could be a result of the high abundance of *Alu* elements in the genome saturating the PCR and limiting specificity for the use of detecting single molecule events.

Figure 2: Alu PCR-fusion band detection by Southern blot



OBJ

Figure 2: Southern blots a) and b) depict various combinations of telomere specific and Alu specific primers used in the current fusion assay. '+ve control' consisted of primers that have detected fusions before in HEK-293 cells*. AF and AR refer to an Alu forward and Alu reverse primer respectively. Arrow depicts putative fusion when using AR and XpYpm primer combinations in a fusion PCR. *XpYpM, 17p6 and 21q1 primers

3.9 LINE-1 fusion assay

3.9.1 Primer optimisation

Primers were designed against sequence data attained for an L-1 element from 'Repbase' at the Genetic Information Research Institute (available online at URL http://girinst.org-repbase). The total size for this LINE was around 5kb. Primers were designed at each end of the sequence (Figure 3) in order to look for fusions in the L1 adjacent DNA. This is important because L1 elements can be quite variable and abundant in the human genome. By looking at the adjacent regions this assay may be more specific for locating fusion points in relation to their position within the human genome.

Each primer combination was tested for the correct product amplification using a temperature gradient PCR. These preliminary data indicated that specific amplification of L1 elements could be obtained (Figures 3 and 4).

The most successful primer (see optimisation figure 4a and b) were then tested in conjunction with those used for telomere fusion PCR (XpYp, 17p and 21q). The detection of putative fusion events was limited (data not shown), therefore to increase the probability of band detection by Southern blotting using probes designed for telomere fusion PCR; L1 specific primers (L3-L6) were added successively to telomere specific fusion primers known to amplify detectable fusion bands. An increase in fusion bands on addition of primers to a control fusion assay would thus indicate additional putative fusion events becoming detectable. These could thus be isolated and characterised by Sanger sequencing.

Figure 3: Location of L-1 primers

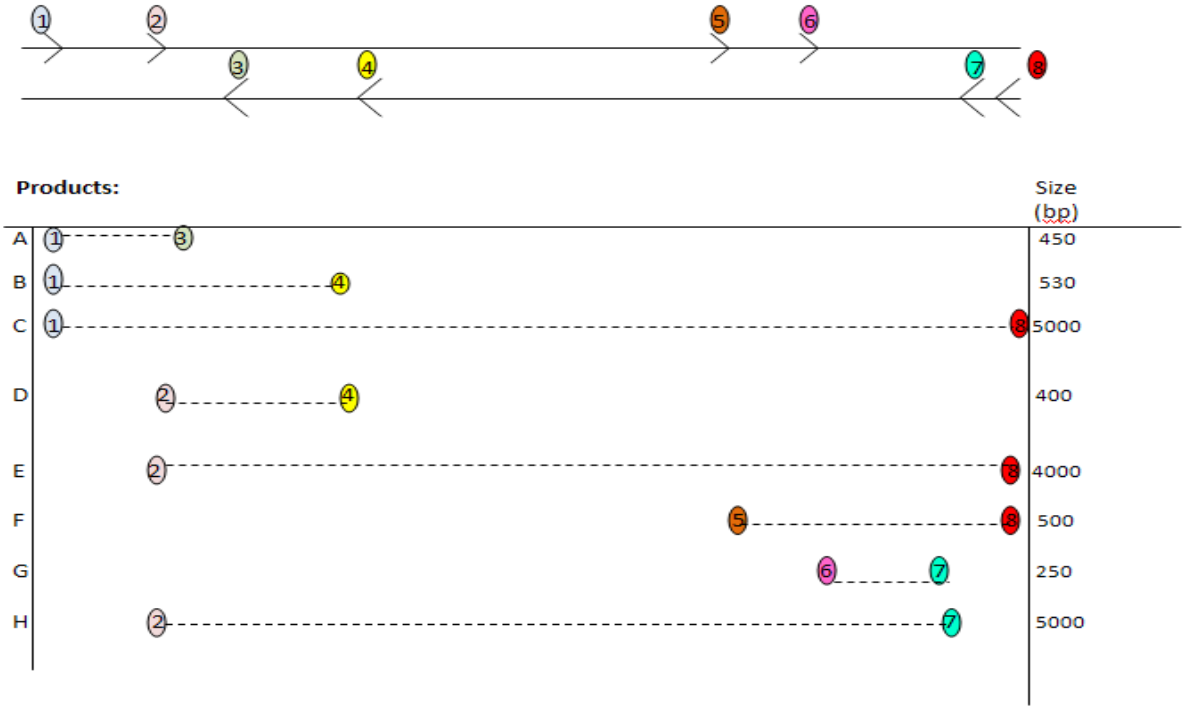
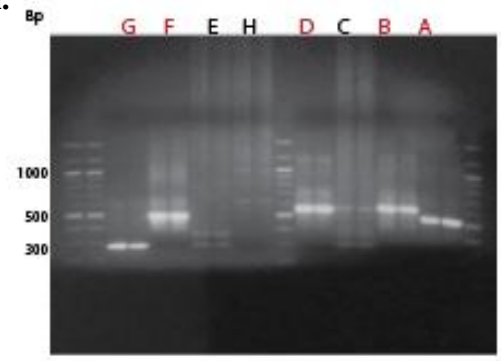


Figure 3: LINE-1 element primers were numbered 1 through to 8 and were thus named L1-L8. Letters A-H on the left hand side of the diagram refers to each primer combination and respective product size generated (sizes of these are on the right hand side (bp)).

Figure 4: Primer optimisation

a.



b.

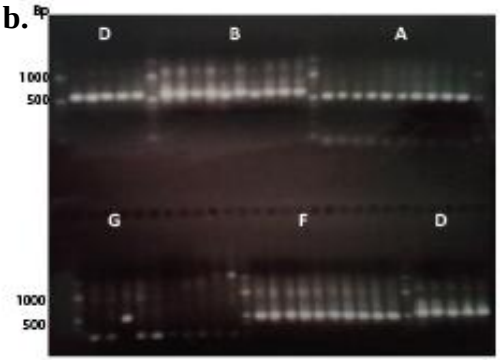


Figure 4a. Shows each primer combination amplified under standard PCR conditions and run on a 1% agarose gel. Letters A-H refer to primer combinations outlined in Figure 3.**Figure 4b**. The brightest product bands (shown in red (4a.)), were then optimised using a temperature gradient PCR.

Figure 5: L-1 PCR inhibits fusion band detection

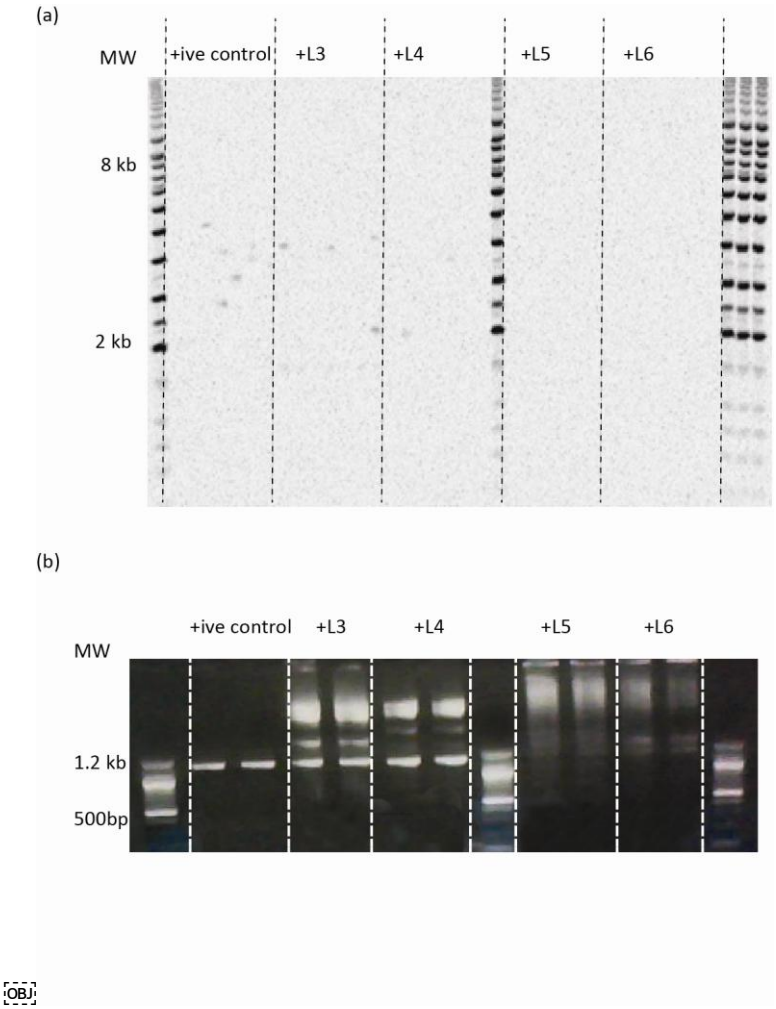


Figure 5: (a) Southern blot: positive control consisted of telomere adjacent primer combinations used successfully previously: $XpYp^{m}$, $17p^{6}$ and $21q^{1}$; LINE-1 (L3-6) primers were added to the control PCR 'mix'. (b) PCR products from the same reactions as (a) resolved on a standard 1% agarose gel without the blotting step

On addition of L-1 primers the quantity of fusion bands detectable on the Southern blot decreased (Figure 5a). This suggested that the fusion reaction was inhibited by the use of

these new primers. This PCR reaction was repeated with additional cycles in the thermocycler (32 cycles), without the Southern blotting step and the products detected on an ethidium bromide stained 1% agarose gel. More product was visible on the addition of the L1 primers into the PCR 'mix'. (Figure 5b); indicating that over-amplification of LI specific products may reduce the single-molecule amplification of the fusion products.

The lack of fusion bands when adding the L-1 primer sets indicated an inhibition of the fusion product. This is the opposite of what was expected and so a better method of detecting rarer fusion types must be developed. Using the current fusion assay in combination with repeat element primers was not a successful strategy for identify additional telomere fusion events.

3.12 Adaptor ligation mediated PCR (ALM-PCR)

Given the apparent lack of success in utilising interspersed repetitive elements to detect telomere fusion products, alternative approaches were tested. Linker driven PCR strategies are used to amplify regions of unknown sequence by 'genome walking'. Genome walking approaches rely on restriction digestion of DNA providing an overhang onto which a linker/adaptor can be ligated to direct PCR synthesis in one direction or another towards or away from a known region to an unknown region. Adapting this technology to find loci fused to telomeres by combining linker technology with pre-designed telomere specific primers, may enable the detection of regions to which telomeres have fused. These assays will have to be adapted to amplify at the single molecule level in order to isolate rare single molecule events.

Several restriction sites near to the XpYp telomere were mapped for three different restriction enzymes (Figure 6). The XpYp telomere is very well characterised and has been used to

develop numerous PCR based assays included, TVR-PCR, STELA and fusion assays due to the unique nature of the subtelomeric regions allowing accurate primer design (Baird 1995; Baird 2003, Capper 2007).

A linker was designed within the XpYp sequence (sequence information generated from Riethman, H et al. URL http://), the reverse strand of which shares 5bp of homology to a 'sticky' *EcoRI* cut site generated during restriction digestion.

3.13 Optimising the linker technology

Control experiments were undertaken to examine if the ligation technology would work. A primer was designed around a specific *EcoRI* site around 2kb from the telomere (Figure 7). This means that a specific band of a known size should be amplified when using two different telomere specific primers. Selections of reverse strand adaptors were designed for optimisation of linker design; a range of sizes were designed because certain protocols using genome walking techniques have found that a shortened adaptor of around 10bp gives greater efficiency. Another factor was to consider end repair priming, which can interfere with the efficiency of the amplification. End repair priming can occur when the overhangs of un-ligated vectors and inserts are filled in and extended by polymerases at the 3′ end during a PCR cycle (Horn et al., 2007). Instead of the ligated insert priming amplification from vector sequences, these un-ligated extended ends can anneal to one another and compete for genuine (i.e. ligation generated) product amplification in the PCR reaction. A phosphate group was thus added to half of the linker sequences (Figure 6b).

Figure 6a: The Position of Various Restriction Digest Sites at The XpYp Telomere

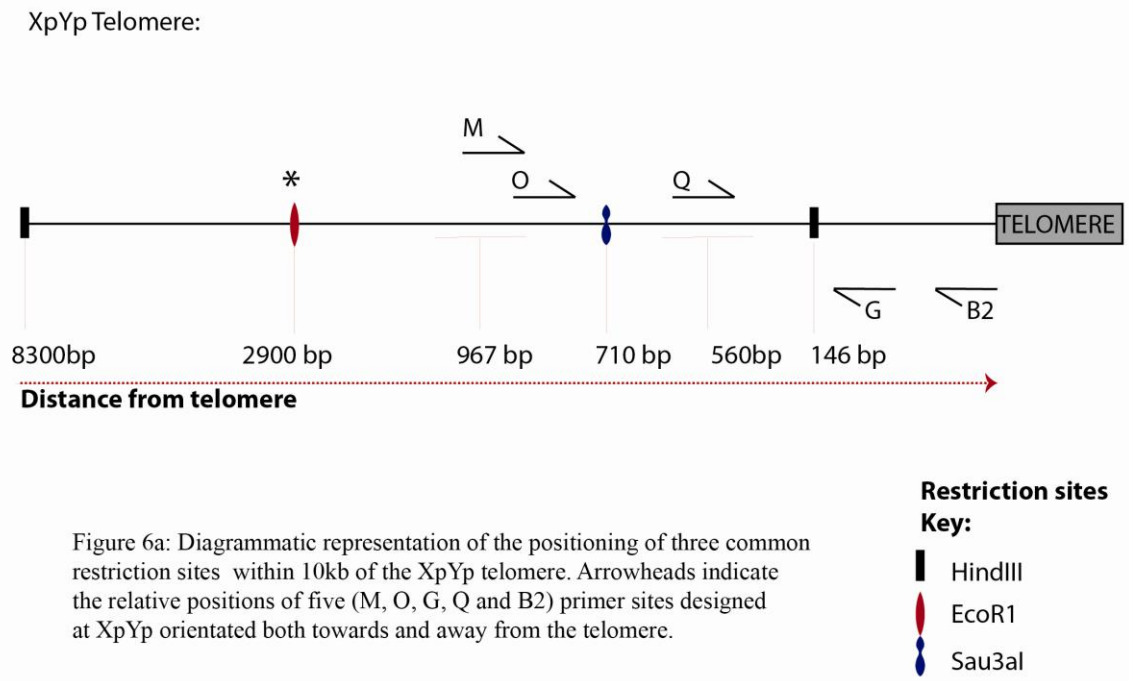


Figure 6a: Diagrammatic representation of the positioning of three common restriction sites within 10kb of the XpYp telomere. Arrowheads indicate the relative positions of five (M, O, G, Q and B2) primer sites designed at XpYp orientated both towards and away from the telomere.

Figure 6b: The Design of Various Oligonucleotide Adaptors at The XpYp EcoRI Cut Site

TGCTCCGTGCATCTGGCATCATG
ACGAGGCACGTAGACCGTAGTACTTAA

*GAATTCCT GGGGACTGCG GATGTAGCTC GCCACAGTAG GTTGTCAGTT- 3′
AACTGACAACCTACTGTGGCGAGCTACATCCGCAGTCCCCAGGAATTC -5′

TELOMERE

XpYp Sequence

EcoRI Linker 1: 5′TGCTCCGTGCATCTGGCATCATG 3′
EcoRI Linker A1: 5′- Phosphate-AATTCATGATGCCAGATGCACGGAGCA 3′
EcoRI Linker A2: 5′ AATTCATGATGCCAGATGCACGGAGCA 3′
EcoRI Linker A3S: 5′ AATTCATGATGCCAG 3′
EcoRI Linker A4S: 5′-Phosphate- AATTCATGATGCCAG 3′
EcoRI PRIMER: 5′ CATCTGGCATCATGAATTCCTG 3′

Figure 6b: A portion of the XpYp sequence containing the EcoRI restriction site (GAATT) 2.9 kb from the telomere is shown in black. Design of various adaptors sharing homology with the cut site were designed (blue) and a linker (red) to anneal to each adaptor forming a double stranded oligonucleotide from which amplification can be directed. The 'PRIMER' consists of both linker and XpYp sequence making the design specific for this cut site locus.

3.14 Testing ligation conditions for an ALM-PCR

The method for ALM-PCR worked by annealing an 'EcoRI' linker with an adaptor linker (Linker A1-A4) to form a double stranded oligonucleotide that will have a portion of homology to the EcoRI cut site (GAATTC) on the lower strand (figure 6b). This linker will be that starting point for the PCR reaction. A primer spanning both the EcoRI linker and a portion of the XpYp sequence at which a known EcoRI restriction site is located was designed for control experiments to test that the ligation of the linker/adaptor was successful. Different ligation conditions were tested with all the different adaptor combinations (EcoRI linker+ Linker A1-A4s- see figure 6b). 'ALM ligation' was used to describe an adapter ligation-mediated PCR method adapted from a previously published method (O' Malley et al., 2007) that used the techniquefor high-throughput mapping of T-DNA inserts in the Arabidopsis genome. 'NEB ligation' utilised the protocol recommended by the manufactures of the T4 DNA ligase (New England Biolabs (NEB)).

3.15 Calculating ligation ratios for ALM-PCR

Ligation ratios were calculated using molar ratio calculations to alter the ratio of oligonucleotide linker to target molecules added to each reaction. Thus ratios of 1:1, 1:3, 1:6, 1:10, 1:50 and 1:100 were calculated accordingly and followed restriction digestion of DNA, these concentrations were used in a series of ligation reactions. The ligations were then inputted into a long-range PCR with a combination of adaptors and a range of XpYp telomere adjacent primers (see Figure 6a and b). Southern blotting using the XpYp o-g probe, used for

telomere fusion PCR assays, allowed the detection of the bands generated (Figure 7.) Ligations termed ‘ALM’ and ‘NEB’ were similar in outcome i.e following Southern blot hybridisation with telomere repeat containing probes both conditions appeared to be visually similar in their ability to amplify bands, and thus NEB ligation was selected for use in subsequent reactions (data not shown). The optimal ligation ratios for EcoRI linker-adaptor combinations using the NEB ligation conditions appeared to be somewhere between 1:3 and 1:10, depending on the primer combination so these conditions were taken into account for subsequent experiments.

A PCR cycle gradient was then used to test the number of cycles required to amplify a single band of the correct size (figure 7b). The PCR cycle number that was concluded to be optimal from this experiment was 28 cycles, due to its ability to amplify the strongest most discernible specific bands, without too much background.

Figure 7a: Ligation ratios: serial dilution of insert:vector

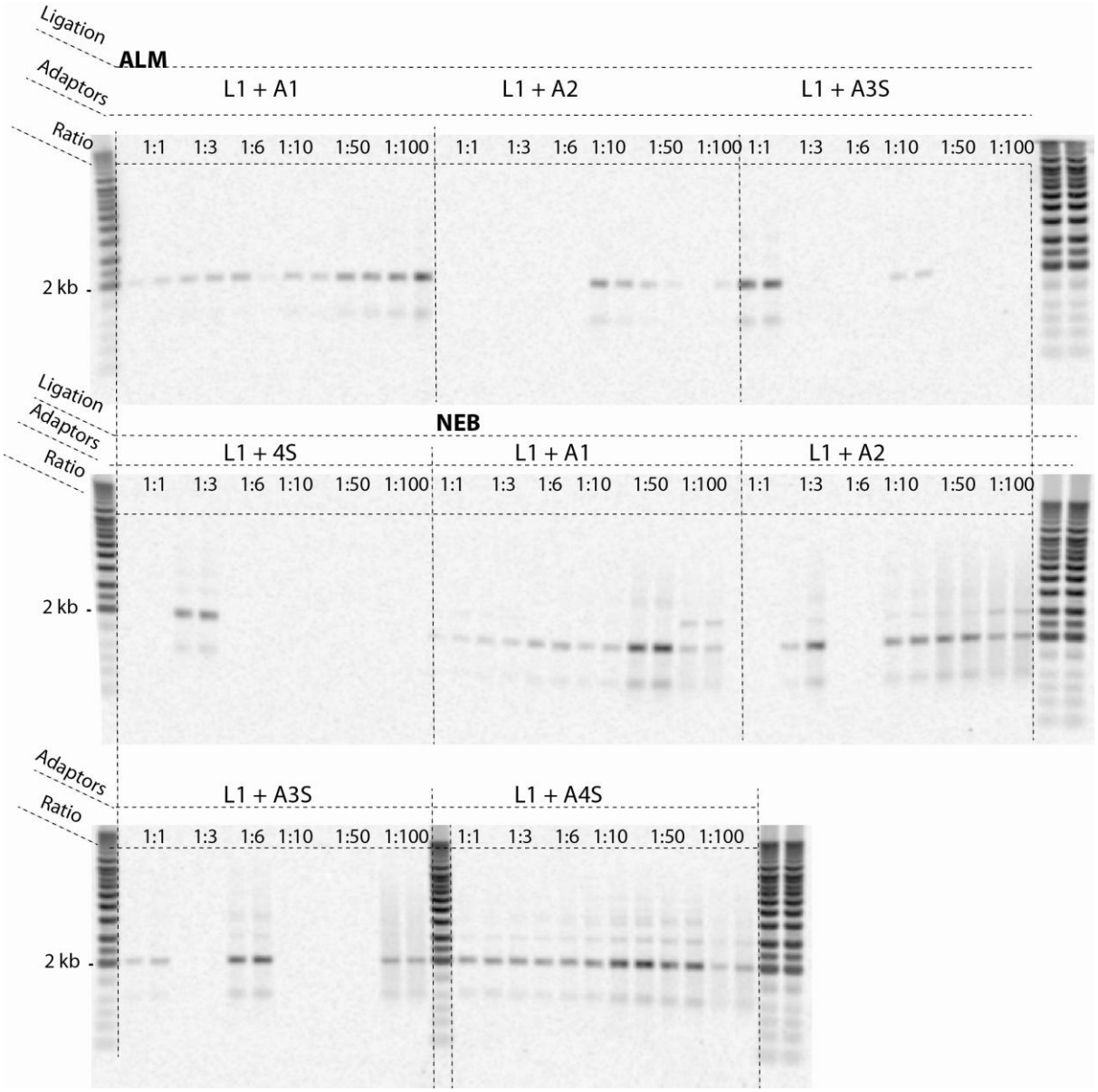


Figure 7a: Ratios of increasing adaptor insert added to digested DNA for ligation reactions are displayed above banding patterns. L1 is the 'EcoRI linker 1' and A1-A4S represent the different adaptors used in combination.

Figure 7b: PCR cycle gradient testing

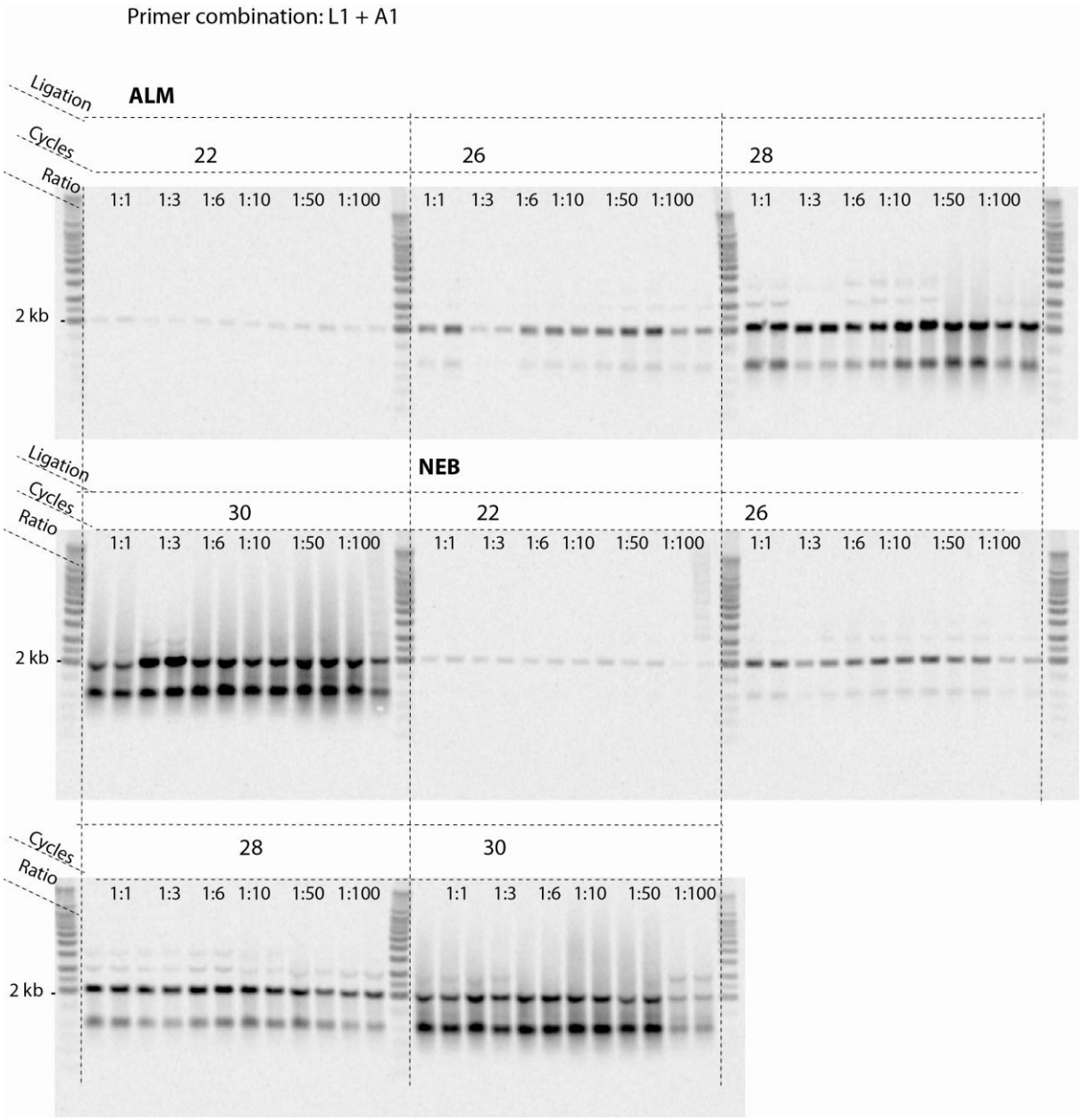


Figure 7b: Different PCR cycles were tested in conjunction with two different ligation conditions plus the variation of ligation ratios. The first four groups (1) were ligated by the NEB protocol and the second set (2) from a protocol by Artandi et al. 2000.

3.16 Titrating to the single molecule level

Once a clean band of the correct control size (roughly 2 kb) was amplified using the selected ligation conditions: NEB ligation reaction with subsequent PCR annealing temperatures of 60ºC (data not shown), 28 PCR cycles and a ligation ratio of 1:6; the reaction was titrated down to the single molecule level (Figure 8). These assays needed to be adapted to amplify at the single molecule, because the importance of this technique is to isolate rare single molecule events occurring in the genome to better understand the telomere dynamics during telomere dysfunction. The number of genomes being amplified in each dilution was calculated from the concentration and moles of input DNA. It was concluded that this assay had the ability to amplify products at the single molecule level and thus single events could be isolated in the genome and characterised. Single molecule efficiency can then be calculated by Poisson distributions to determine the probability or rarity of the fusion event occurring and we could thus test this by seeding reactions into undigested DNA.

Figure 8: Titrating to the single molecule level

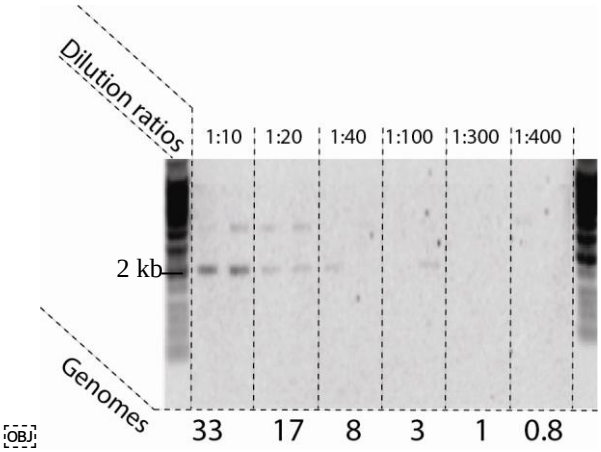
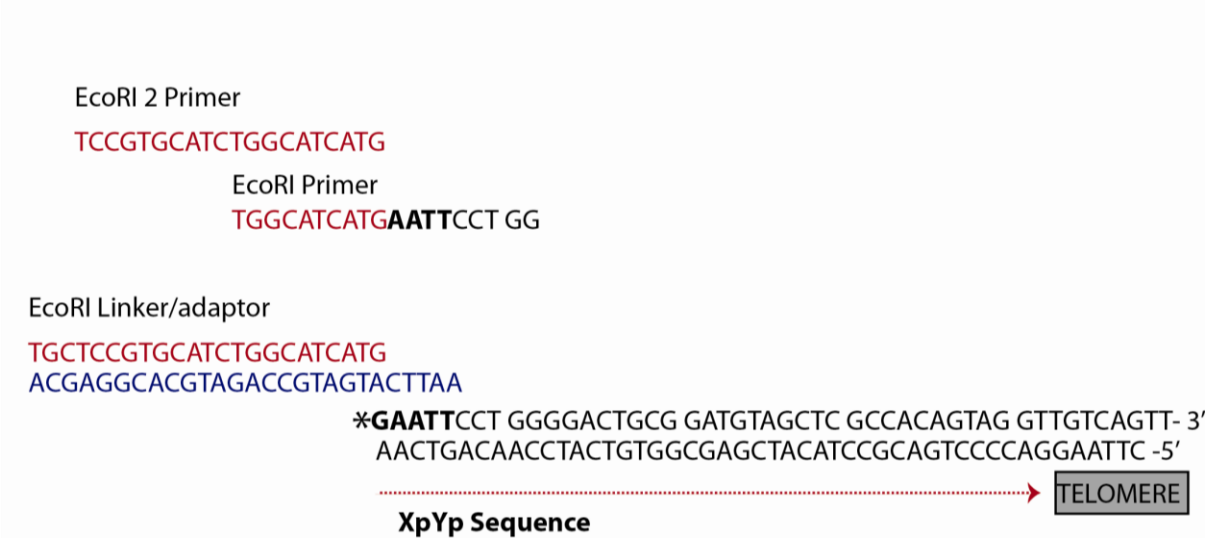


Figure 8: Bands representative of specific amplification appear sporadic when diluting the DNA to the single molecule level using ligation conditions highlighted in section 3.16- ALM ligation NEB ligation and annealing temperature of 60ºC. Numbers shown at the bottom of the blot are representative of the number of genomes being amplified.

Figure 9: New primer design



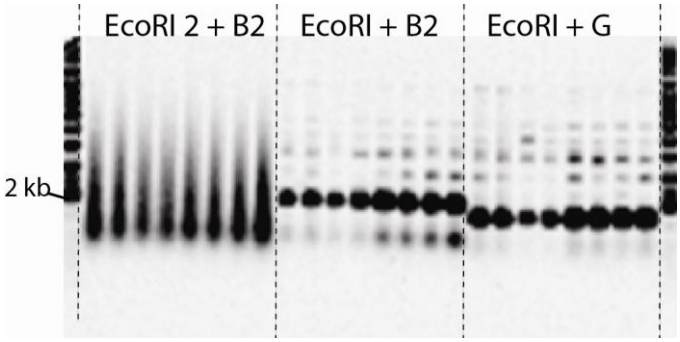


Figure 9: A primer was designed that was specific to the linker region only (blue letters). The result from using this primer is shown above in a southern blot. ECORI2 is the newly designed primer and the PCR reaction product was visualised by Southern blotting (First 8 reactions on blot; bottom). Alongside this, the previously utilised ECORI primer was used in a PCR reaction with XpYp primers B2 and G (using optimum ligation conditions discussed in section 3.16 and an annealing temperature of 60C). Genome equivalents represented here are 33 (1:10 dilution).

3.17 Testing linker specificity

As mentioned before (3.14) a primer was designed to span the linker and a known region at XpYp (where a cut site is located). This was useful to test the success of the ligation. A linker primer was hence designed that would only be homologous to the linker. Thus when the portion of DNA containing the annealed linker is synthesised, this new primer will be able to bind to initiate synthesis between the linker and telomere adjacent primers.

When using this new primer design (EcoRI2) in conjunction with XpYpB2, large smears were visible on the gel, differing in appearance from the control bands adjacent. Diluting out the PCR product and seeding it into new PCR reaction was done in the attempt to clean up

the PCR reaction and thus isolate a clear band as it was hypothesised that PCR artefacts may be interfering with the production of bands akin to fusion events. This approach failed to amplify anything but smears. This may be evidence that too many products from the rest of the genome were being amplified. (Figure 9) or that the products generated are non-specific.

3.18 Conclusions:

High ligation efficiency was achieved by the various optimisation steps for this technique. Control experiments were thus successful for this purpose. The *EcoRI* 2 primer products were only visible as smears on the gel that are consistent with the amplification of other products in the genome putatively crowding the PCR reaction. This may show that the design of linkers were not specific enough for fusion amplification. Titrating the DNA has shown that this ligation technique can amplify events at single molecule level. Thus despite many attempts and alternations of ligation and cycling conditions it was not possible to obtain evidence that was consistent with the amplification of telomere fusion events.

3.19 A High Throughput Splinkerette PCR

A genome walking protocol used to screen for transposon inserts within the genome of mice was adapted for the purpose of telomere fusion detection. This is due to the inability of the previous assay to amplify products specific enough for detection and isolation. Interference of putative artefacts in the PCR/non-specific products made the detection of any fusion events too difficult. It was thus deduced that the linker used for amplification needed more specificity to ensure correct amplification of telomere related fusions. The first principle of the technique used is to digest the DNA with a *Sau3aI* restriction enzyme. To this a

specialised ‘splinkerette’ adaptor can be annealed as a template to drive PCR from these restriction sites. The splinkerette consists of annealed oligonucleotide pairs. The annealing of pairs of oligonucleotide adaptors creates an overhang compatible with overhangs generated by restriction of the DNA with *Sau3aI* enzyme. Primers to drive PCR synthesis are homologous to the splinkerette sequence (not complementary) and can thus only bind to newly synthesised strand of splinkerette adaptor. This increases the specificity of the PCR amplification reaction. The splinkerette adaptor design is gives the reaction a higher degree of specificity compared with the adaptors designed previously (see ALM-PCR), this is due to the presence of a hairpin structure that is incorporated on the lower strand, which prevents end repair priming (Figure 10).

Figure 10: Structure of a Splinkerette

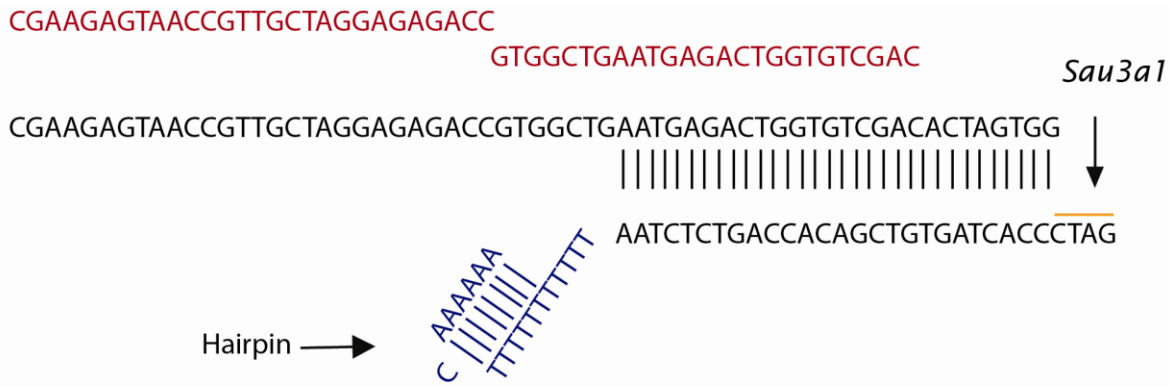


Figure 10: The splinkerette structure has a long strand and a short strand to which the ‘Splink’ adaptors can anneal. The two homologous sequences in red are the 5’ to 3’ sequences for primers Splink 1 (top left) and Splink 2 (bottom right). The A-T rich Hairpin structure prevents end repair priming.

3.20 Testing ligation efficiency

As with previous genome walking experiments it was necessary to design control experiments that would confirm successful ligation and amplification of a known product size between at the *Sau3aI* restriction site at XpYp. A primer (SP1) was designed that overlapped both the splinkerette adaptor sequence and the sequence that it would have to anneal to at XpYp (restriction site). Figure 11 shows the design for this SP1 primer encompassing both the adaptor and the XpYp sequence around the cut site. This primer will be used alongside a predesigned primer (Splink 1) Uren.A, 2009); Splink 1 is not specific to XpYp and is homologous to the splinkerette sequence. (Map of GATC cut site and orientation of XpYp primers is shown in figure 7a).

Figure 11

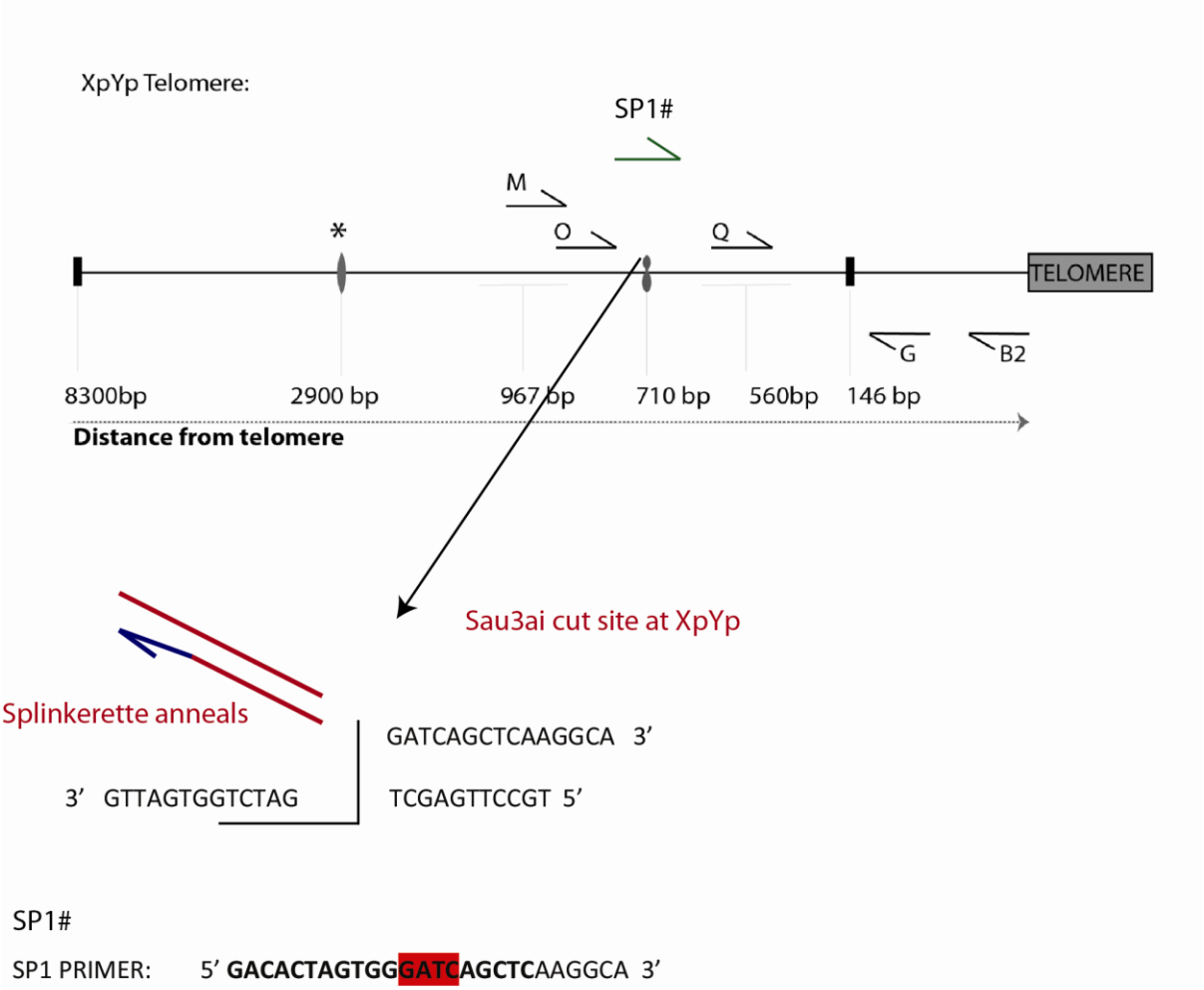


Figure 11

Figure 11: An illustration showing the Sau3aI cut site at XpYp and the way in which the Splinkerette (double stranded-shown in red with blue hairpin) will anneal to the GATC sequence. At the bottom is the sequence of an SP1 primer that spans the Splinkerette sequence and the XpYp sequence (shown in relation to XpYp at the top, see SP1#).

3.21 : Optimisation of the Splinkerette ligation

Two different ligation conditions were applied with the new splinkerette adaptor mix and the digested 293 DNA (as before, see 3.12) and tested using the SP1 primer. These ligations were optimised for annealing temperature at which the primers work most efficiently by temperature gradient PCR. Resulting blots showed that a temperature of between 60 and 65 degrees was optimum and the ligation conditions showed no difference in efficiency, thus the NEB protocol was used henceforth (data not shown).

3.22 Results: Preliminary data for ligation of Splinkerette

The standard Splinkerette protocol outlined by Uren and colleagues (Uren et al., 2009) using the Splink -1 (non-specific) primer, was tested in a nested PCR with the XpYp specific primers (XpYpM and XpYpO). Several bands were visible on the Southern blot following the second round of amplification with XpYpO, one of approximately 970bp was consistent with the size of fragment expected following amplification from Splinkerette linker ligated to the XpYp Sau3AI site. Different volumes of PCR product were run on the agarose gel, due to the presence of large smears in preliminary experiments and by adding less product the control bands became clearer (figure 12a). In the presence of a successful ligation of the Splinkerette to the Sau3AI site, it should be possible to successfully amplify products of defined sizes between the Splinkerette and XpYp specific primers on either side of the Sau3A1 site. These products serve as an internal control for the ligation and amplification stages. To further verify the specificity of the amplification further control experiments using the SP1 and the Splink 1 and 2 primers (Figure 11 a) were used in combination of the XpYp primers on either side of the Sau3A1 site (Figure 12b-c). Multiple control bands of the expected sizes depending on the XpYp specific primer used in the PCR were amplified. No product was

amplified in the absence of ligase. Together these indicated successful and specfic ligation as well as the correct annealing of primers to the splinkerette ligation site.

Figure 12: Optimising the Splinkerette PCR assay using XpYp primer combinations

12a:

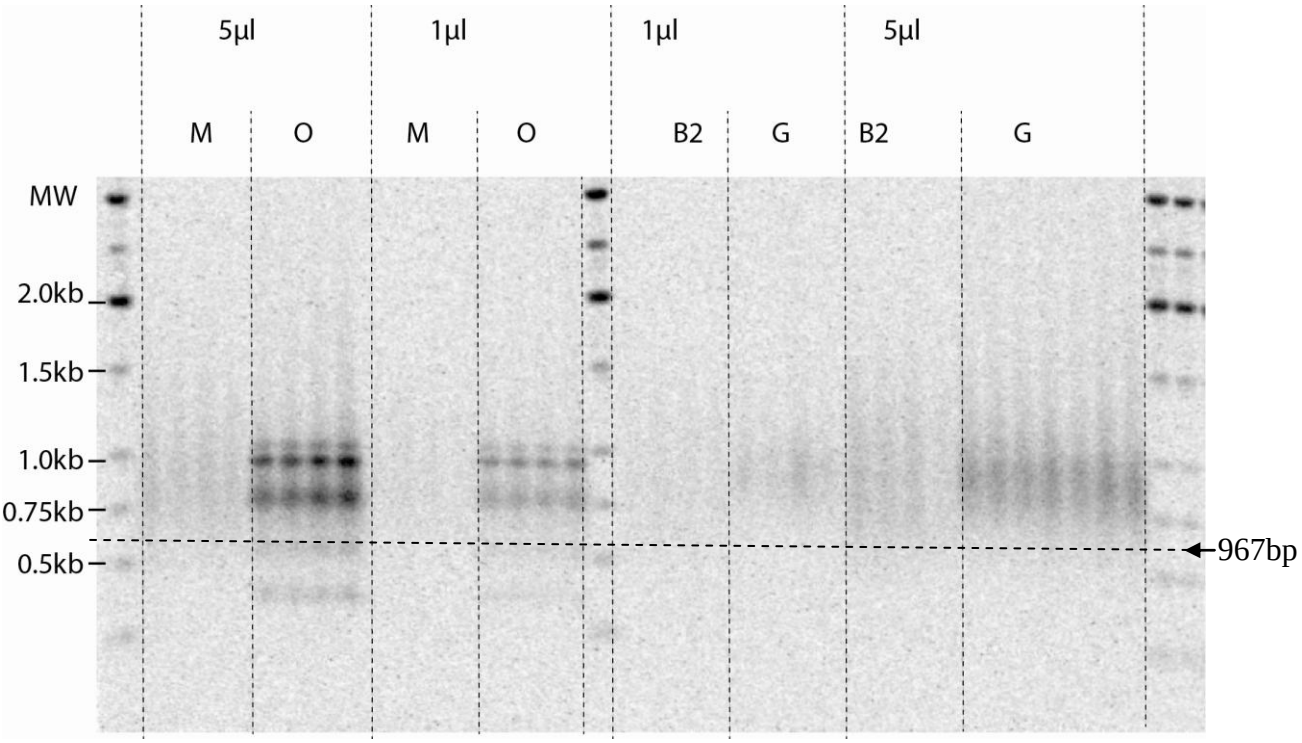
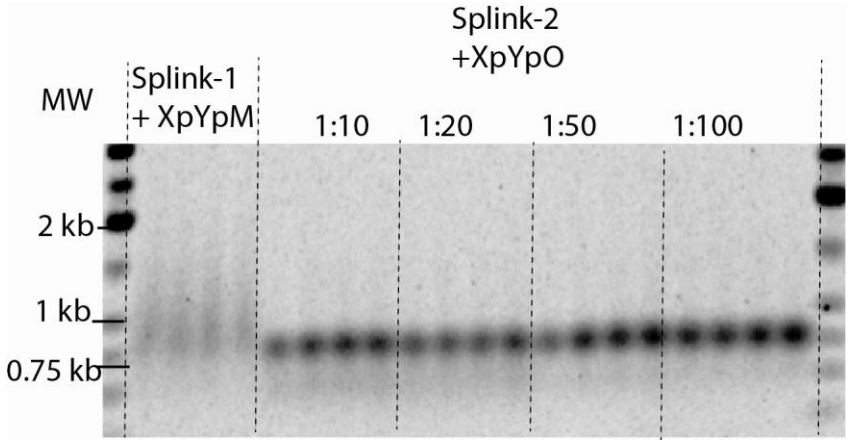


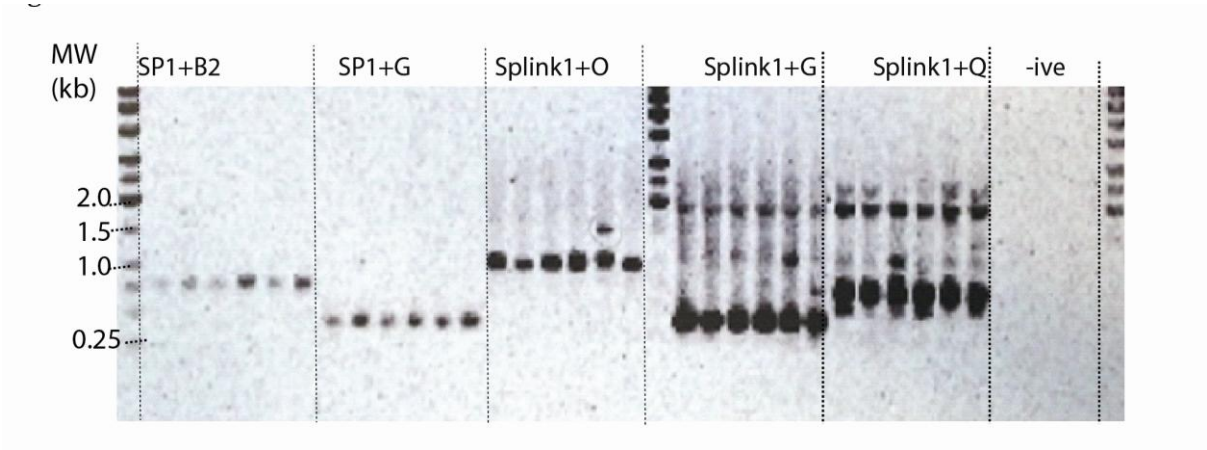
Figure 12a: Splink primers were used in combinations in a nested PCR, for example Splink-1 and XpYpM were used in the first round of PCR and then Splink-2 was used with XpYpO (shown at the top) in the second round. Different volumes of product were run on the gel (5ul and 1ul) in an attempt to improve resolution.

12b)



12b: Nested PCR: To remove any interfering products the first round of PCR product (Splink1 +XpYpM) was diluted to 1:10, 1:20, 1:50 and 1:100 in distilled water before being used in a second round of PCR (Splink2 +XpYpO).

12c)



12c) The SP1 primer was used in a fusion PCR assay with a combination of XpYp primers (B2, G, O and Q). The negative control consisted of primers Splink1 and XpYpO using DNA that had not undergone digestion or ligation with Splinkerette adaptors (i.e. this was a negative ligation control).

3.23 Single round of PCR

Given the apparent success and specificity of the splinkerette ligation rations, further experiments were undertaken to establish if it was possible to detect products that were consistent with the amplification of telomere fusion. Using a new ligation of digested Hek-293 DNA with the splinkerette (Figure 13) and a single round of PCR instead of a nested PCR. The resulting blot showed multiple bands for all primer combinations together with the constant bands serving as a positive control for ligation to the XpYp Sau3AI site.

The SP1 primer was initially designed as a 'control' primer i.e. to test the specificity of splinkerette ligation in its ability to amplify products with known XpYp telomere adjacent primers. As mentioned previously the design enabled this by designing half of the oligonucleotide sequence to share homology with the splinkerette linker and the other half with a stretch of XpYp sequence occurring immediately after the Sau3ai restriction site. The splinkerette could potentially anneal to upper or lower DNA strands during the ligation process and thus amplification of PCR products when using XpYp adjacent primers is possible from either side of the GAATC cut site. For this experiment it would be expected that bands cold be amplified using an array of different XpYp-Sp1 primer combinations, however they would be of a known size, i.e. the position of both the splinkerette annealing site and the XpYp primers are known.

Interestingly the resulting blot showed not only the expected constant product band of predetermined size, but also an array of multiple bands above and below this seemingly constant band for all primer combinations. The banding pattern obtained contained randomly sized bands consistent with single molecule amplification, was similar to that observed in telomere fusion reactions using the original technology. Due to the orientation of SP1 being towards the telomere, these bands could be evidence of fusion events between XpYp and other loci being picked up. Alternatively these bands could be artefacts from the PCR reaction. To verify these bands PCR product was seeded into a reamplification PCR with an

increase amount of PCR cycles, in the hope that these bands could be isolated for Sanger sequencing. This approach however proved unsuccessful, when running the reamplified PCR products on an agarose gel products appeared as smears under Ultraviolet light (data not shown).

Figure 13: Splinkerette PCR was performed on a new DNA ligation with a Single round of PCR

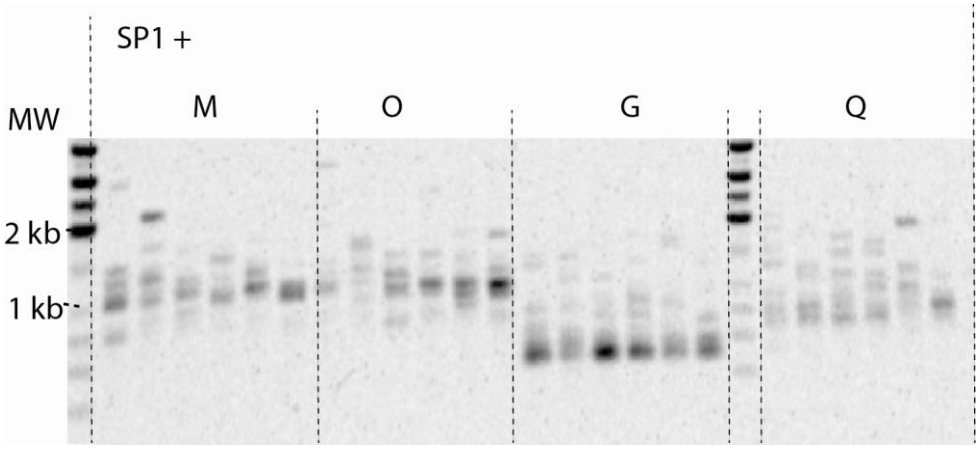


Figure 13: The SP1 primer was used in a fusion PCR assay with a combination of XpYp primers for just one round of PCR instead of two.

3.26 Elucidating possible telomere fusions

In order to isolate fusion events between *Sau3aI* sites in the genome and telomeres, the PCR reactions were repeated and diluted to the single molecule level (based on starting concentrations of DNA at 40ng/ul). This was done in the hope of characterising any rare single molecule fusion events occurring in the genome.

First a diltution series of Splinkerette ligated DNA between 1:10 and 1:100 was used in a PCR reaction in the hope of single molecule level amplification. Genome equivalents being amplified using the most dilute DNA (1:100) was estimated to be 3, assuming starting DNA

concentrations of 40ng/ul after successful digestion with the Sau3ai enzyme. However the initial dilutions produced limited detectable bands when hybridised with an XpYpo-g probe for anything after 1:10. It was thus proposed that starting DNA concentrations could have been overestimated, and thus a dilution series between 1:1 and 1:10 was used for a second set of PCR reactions using the same primer pairs. The resulting blots show multiple sporadic bands (Figure 14) reminiscent of fusion events seen with the pre-existing assay for fusions between telomere groups.

Figure 14

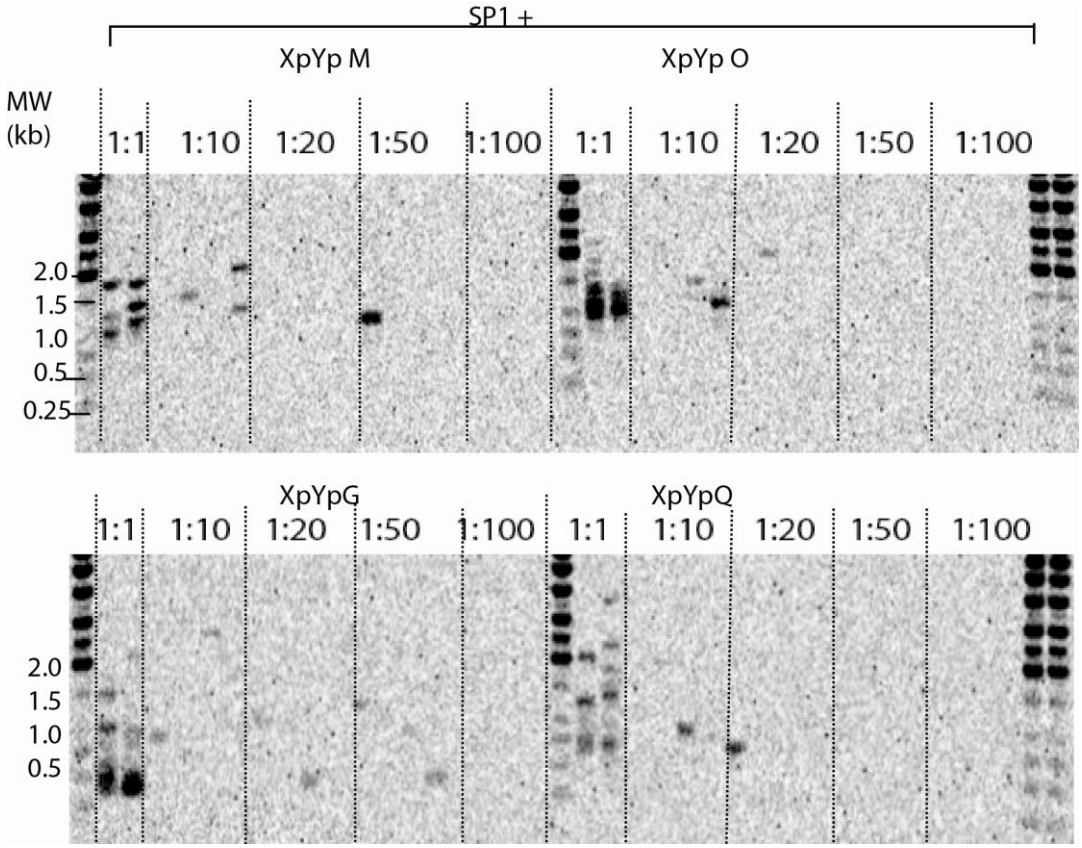


Figure 14: The input Splinkerette ligated DNA was diluted down in a series of different dilutions ranging from 1:1 to 1:100. These dilututions were then used in a PCR reaction with the XpYp telomere adjacent primers used previously: XpYp^{m} and XpYp^{o} orientated towards the telomere; and XpYp^{g} and XpYp^{q} orientated away from the telomere. Genome equivalents amplified are estimated to be about 33 for a dilution of 1:10 and 3 for a dilution of 1:100.

Figure 15:

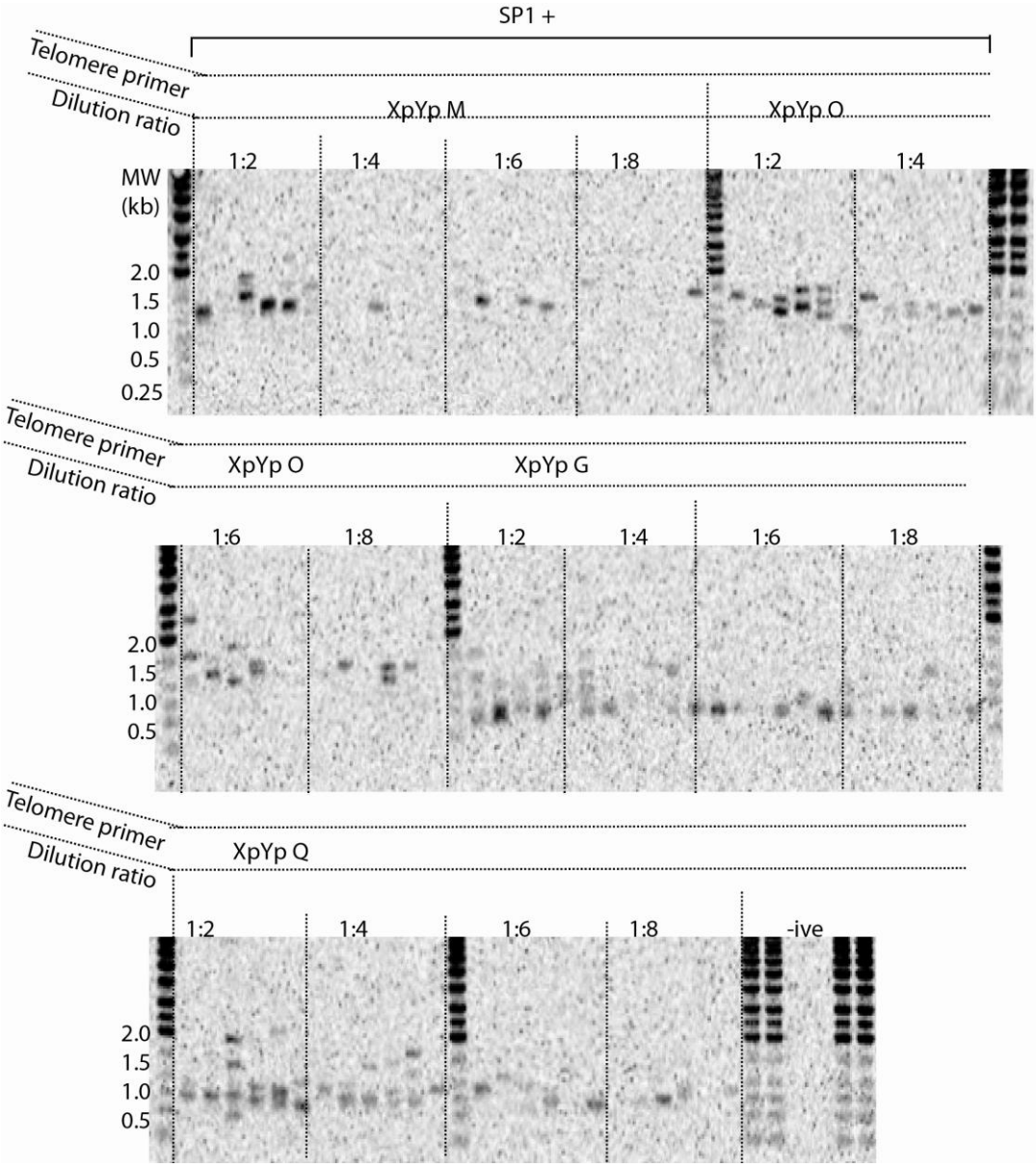


Figure 15: The dilution of input DNA into the splinkerette PCR was expanded between 1:1 and 1:10 to try to increase the yield of bands with the potential to reamplify and sequence events.

3.27 Results: Sequencing fusions

A putative fusion band was isolated and sequenced in the hope that a rare single molecule fusion event involving a telomere could be characterised when using this Splinkerette technology. Unfortunately the only successful attempt at sequencing an 'event' of this type revealed a portion of sequence at the XpYp telomere between the SP1 primer annealing site (Sau3ai site) and the $XpYp^{Q}$ primer (See figure 16). Although this is not consistent with a telomere fusion event, this does further confirm that the Splinkerette has ligated successfully to the Sau3ai site due to the correct amplification between the SP1 primer and XpYp telomere adjacent primer. This concludes that this Splinkerette assay has potential for screening fusion events in the genome.

Figure 16: Sequenced junction at XpYp

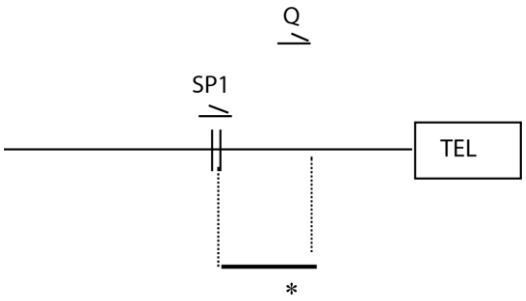


Figure 16: Splinkerette PCR with XpYpQ (Q) and SP1 primers isolated a band of just under 600bp. This region when sequenced was characterised as the region at XpYp between the Sau3ai site and the XpYpQ primer (illustrated by the thick black line* on diagram).

3.28 Design of a new SP2 primers

The banding patterns generated using the SP1 primer in combinations with XpYp primers were reminiscent of similar banding patterns achieved using the current telomere-telomere fusion assay. However the design of SP1 was intended for use as a control experiment to optimise the ligation reaction at the single molecule level and in this, regardless of fusion-like bands, failed in its ability to characterise genuine fusion events. New primers were designed

with decreasing specificity to the XpYp known region. Subsets of primers were also designed in the opposite orientation of the SP1 primer pointing away from the telomere. Four primers were designed, each with decreasing specificity to the XpYp restriction site i.e. number 1 had the most homology to the XpYp telomere sequence and number 4 no homology to the XpYp site and was purely specific to the splinkerette linker. (Figure 17) It was speculated that the fourth, least homologous primer might be able to detect fusions involving telomeres that had annealed elsewhere in the genome in proximity to other Sau3ai sites.

Figure 17: Primer design

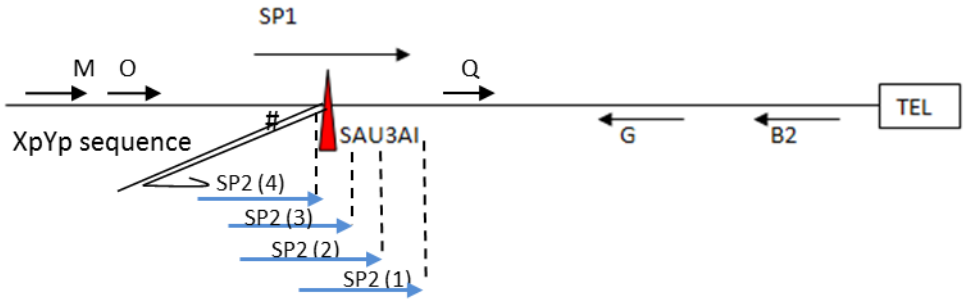


Figure 17: Diagrams show the design and orientation of the SP2 primer sets (blue arrows), the sequence of each different SP2 primer has varying specificity to the XpYp sequence adjacent to the known Sau3ai cut site (red triangle) i.e. SP2(1) is similar to SP1 and has an overlap sharing homology to the XpYp sequence; whereas SP2 (4) is only specific to the splinkerette linker (illustrated as double stranded region (#)). Dashed lines illustrate various degrees of overlap each primer has with the linker (Splinkerette) and XpYp sequence. Black arrows show the simplified orientation of the XpYp primers to be used in experiments in combination with the SP2 primer sets.

Figure 18: Fusion PCR using new primer combinations

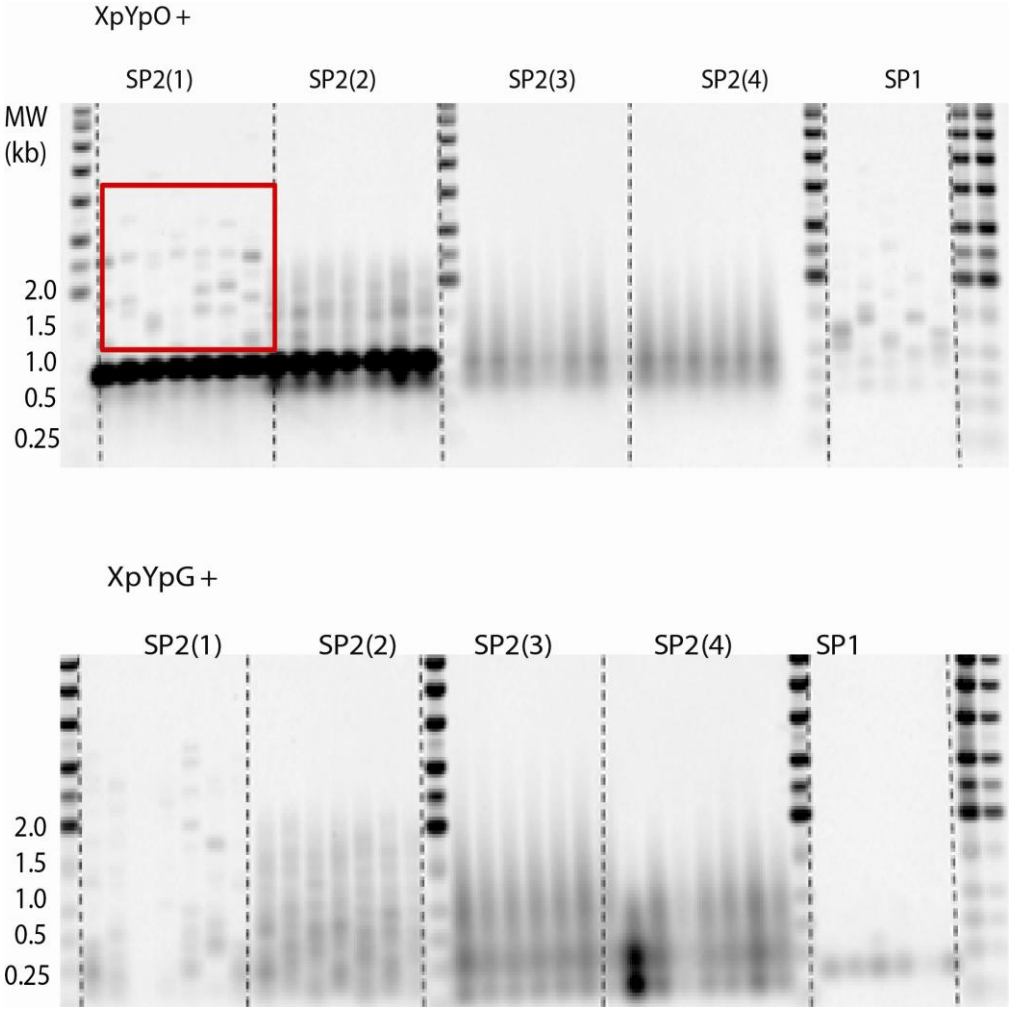


Figure 18: Southern blot showing splinkerette PCR results using a combination of XpYpO and G with the four different SP2 primers. SP1 with the XpYp primers was used as a positive control. . The area above the constant band in the reaction including SP2(1) and XpYpO (outlined in red) was cut out from the gel to be sequenced

3.29 Results: Testing out new SP2 primers designed away from the telomere end:

Each SP2 primer (1-4) has successively less specificity to the XpYp sequence and slightly more to the splinkerette sequence. This gradual loss of specificity results in less clear banding patterns and more smearing. This could be due to the vast amount of product being amplified.

Dilution of the DNA input had no effect on the smearing of the bands. The area above the constant band in the reaction including SP2(1) and XpYpO (outlined in red, figure 18) was cut out from the gel to be sequenced, however this was unsuccessful. Regardless of this it is the first time that this type of fusion-like banding pattern has been obtained using primers in the correct orientation, i.e. pointing away from the telomere. Time constraints with this work impeded the ability to continue with this technique to characterise exactly what these bands akin to fusion events could be, however with more work this assay could be utilised as a potential way to screen for fusions involving telomeres in the genome.

3.32: Conclusions of the splinkerette PCR

Whilst banding patterns that were consistent with the detection of fusion could be obtained using the splinkeretter technology, with the exception of one event it was not possible to systematically isolate and characterise these fusion events. Due to time limitations this aspect of the project was not pursued further.

3.33: Discussion

3.33.1: Fusions between repetitive element and telomeric loci

Finding fusions by utilising primers designed against repeat elements found in the human genome proved limiting. *Alu* and L1 elements are highly abundant in the human genome (Graham and Boissinot, 2006) and this could explain why results were limiting i.e. overcrowding of the PCR reaction could inhibit or mask the detection of any specific single molecule fusion events involving telomeres.

3.33.2: Adaptor ligation mediated PCR for a novel fusion assay

Ligation-mediated based PCR strategies have shown a relative amount of success in regard to genome walking, however, the design of the linker/adaptor sequences plays a key role. If the linker anneals to multiple non-specific sites then the detection of specific sites will be unsuccessful. End-repair priming is a drawback of these techniques. End-repair priming involves the free sticky ends of un-ligated linkers which are based on restriction enzyme sites that generate 5' overhangs. These ends are filled in during the first round of the PCR reaction and are thus able to anneal to one another with sufficient stability to initiate priming (Devon et al., 1995; Horn et al., 2007). This leads to exponential PCR amplification, which could explain any excess non-specific product on Southern blots. Dimers, i.e. linkers annealed to each other, can also be generated by end-repair priming and the two halves will form a hairpin (Uren et al., 2009), this could be detrimental to product amplification and could explain why L1 technology showed inhibition of PCR product. This is why a splinkerette linker was adapted for use in subsequent experiments, to increase specificity of amplification.

3.33.3: Splinkerette PCR fusion success

The results generated from the splinkerette assay were extremely dependant on the successful ligation of the splinkerette itself, a number of negative results were generated in preliminary experiments before the ligation was successful, highlighting the importance of optimisation.

The Splinkerette PCR provided the most promising data, however there was limited success for sequencing any putative fusion events involving telomeres. The single fusion sequence obtained was not sufficient to deduce that the technique was successful. It is possible that the splinkerette technique was capable of detecting telomere fusion events but the limitation might have been in the difficulty of re-amplify the products for sequencing. Thus future directions could involve applying a next generation sequencing approach, such as Illumina sequencing, to try to ascertain what is being generated in the splinkerette PCR amplification without recourse to re-amplification of the fusion products. NGS would allow the full spectrum of fusion events occurring in the genome-wide setting to be sequenced; this technique has been used successfully for the sequencing of whole genomes in parallel to compare mutational profiles in breast cancer (Reis-Filho, 2009). This technique at the moment is not successful in identifying novel fusion events within the genome.

Chapter 4: Telomere dynamics in Breast Cancer

4.1: Abstract

Telomeres are structures that cap the ends of chromosomes, they prevent the natural end of a chromosome from being recognised as a double-stranded DNA break. Dysfunctional telomeres may either trigger replicative senescence, or undergo fusion with other chromosome ends or non-telomeric DNA breaks. Fusion can result in the formation of dicentric chromosomes and the initiation of cycles of anaphase-bridging, breakage and fusion, which in turn can cause large-scale, potentially oncogenic, genomic rearrangements such as non-reciprocal translocations.

This project aims to test the hypothesis that telomere dysfunction occurs during the progression of breast cancer and that this can drive the large-scale genomic rearrangements frequently observed in this disease. To do this, tumour derived DNAs from a cohort of 130 patients with invasive ductal carcinoma of the breast were analysed for telomere length using single telomere length analysis (STELA). STELA is a high resolution single-molecule telomere length analysis strategy that allows the full spectrum of telomere length to be detected, in particular this includes short telomeres within the length ranges at which telomere dysfunction and fusion can occur.

STELA was used to analyse three telomeres in invasive ductal carcinoma (IDC): XpYp, 17p and 2p. This has allowed the high resolution characterisation of an array of telomere dynamics in IDC including critically short telomeres and LOH.

4.2: Background to experiments: Telomere dynamics in breast carcinomas

4.2.1: Telomere length studies in epithelial and mammary cancers in mice

A key gene that is often mutated in breast cancer is the p53 gene (Gasco et al., 2002). P53 has many tumour suppressive mechanisms, with roles in apoptosis, genomic stability, and inhibition of angiogenesis (Schmitt et al., 2002). P53 can activate DNA repair proteins; can arrest cell growth at the G_1/S cell cycle point; or initiate apoptosis if DNA is damaged beyond the point by which it can be repaired by enzymatic machinery (Schmitt et al., 2002; Gasco et al., 2003).

Previous observations have shown that the loss of p53 in cells with telomere dysfunction increases cell survival, which confers a greater oncogenic risk (Artandi et al., 2000, Blasco et al., 2011). A study by Artandi *et al.* showed that in mTERC$^{-/-}$ (telomerase RNA component knockout) mice, continuous telomere shortening led to an increase in the detection of phenotypic characteristics of genomic instability such as aneuploidy and NRTs in ageing mice (Artandi et al., 2000). Telomere fusion events (observed by *FISH*) resulted in higher rates of multiple types of cancer tumour formation (Artandi et al., 2000). In this study mice were grown with successively deficient telomerase components (mTERC$^{+/+}$, mTERC$^{+/-}$, mTERC$^{-/-}$) with mutated copies of the gene encoding p53 (Artandi et al., 2000). In later generations of mTERC$^{-/-}$ mice, decreased telomere function was observed and this correlated with earlier onset of tumorigenesis (multiple adenomas and mammary cancers) (Artandi et al., 2000). Short dysfunctional telomeres were present in p53$^{+/-}$mutants, which correlated with earlier tumour onset (Artandi et al., 2000). This evidence shows that dysfunctional telomeres combined with a p53 deficiency may contribute to early onset of tumorigenesis.

The mechanism behind mTERC$^{-/-}$ p53$^{-/-}$ mouse telomeres becoming dysfunctional and leading to accelerated tumourigenesis was studied using *FISH* in mouse mammary cancer tumours (Artandi et al., 2000). Telomere PNA fluorescent *in situ* hybridization (FISH) was

carried out on metaphase spreads of breast cancers cells attained from mTERC-/- mice (Artandi et al., 2000). Signal-free ends (no telomere signal) were observed in dysfunctional telomeres, along with end-to-end fusions and a high number of non-reciprocal translocations (NRTs) (Artandi et al., 2000). It was proposed in this study that NRTs arise via dicentric chromosome formation due to telomere dysfunction (Artandi et al., 2000). Thus telomere instability and shortening may contribute to the initiation of breast neoplasia.

4.2.2: Telomere length in human breast tissue

Consistent with the majority of tumour types, telomere erosion has been demonstrated in breast cancer. In a study by Heaphy *et al.* telomere DNA content was measured using a slot blot titration assay. This assay gives a mean percentage of telomeric DNA concentration by calculating the ratio of telomeric DNA content to centromeric DNA content. In paraffin embedded invasive human breast cancer samples, tumours with the shortest telomeres (lowest telomere DNA content) showed an increase in detectable aneuploidy by flow cytometry. Human breast tumours with the shortest telomere length had the greatest amount of metastasis (significance $p < 0.05$) (Heaphy et al., 2007).

A telomere-driven crisis has been considered to play a role in the development of breast carcinoma, in particular during transition from UDH (usual ductal hyperplasia) to DCIS (ductal carcinoma *in situ*) (Chin, K. et al., 2004). Telomere crisis is characterised by telomeres that have lost their ability to cap the chromosomal ends (Greenberg, 2005). The relation between genomic instability and telomere length was observed by Chin *et al.*, using *FISH* in normal ductal epithelium, UDH, DCIS and invasive cancer (Chin, K. et al., 2004). Genomic instability was assessed phenotypically by confocal microscopy analysis of tissue sections. And an increase in anaphase bridges was detectable in UDH and DCIS (Chin, K. et al., 2004). Telomere length analysis by *FISH* showed telomere length was significantly shorter in invasive cancers than DCIS and shorter in DCIS than UDH (Chin, K. et al., 2004).

There was thus a correlation was shown between genome instability, telomere length and progression to invasive breast cancer.

One study by Heaphy *et al*. focussed on evaluating telomere lengths in different cancer subtypes based on tumour characteristics (Heaphy et al., 2010). Using *FISH* at the single cell level it was found that telomere lengths were shorter in more aggressive subtypes, such as luminal B, HER-2-positive and triple-negative breast tumours, suggesting that tumour telomere length could be an important tool as a prognostic marker for breast cancer (Heaphy et al., 2010).

4.2.3: DNA repair, Telomere length and breast cancer: activation of ATM

DNA damage has been revealed to take place as an early event in the majority of human epithelial cancers and many other types of cancer (Kirsch et al., 1998). Studies have shown that abrogating genes involved in DNA damage repair pathways, such as p53, results in an increase in cancerous cells/tumours *in vitro* and *in vivo* (Kirsch et al., 1998). Ataxia telangiectasia mutated (ATM) is a protein kinase activated in response to DNA double-strand breaks (Shiloh, 2006; Lee and Paull, 2007). It phosphorylates several key proteins that initiate activation of the DNA damage checkpoint, leading to cell cycle arrest, DNA repair or apoptosis. Several of these targets, including p53, CHK2 and H2AX are tumour suppressors (Shiloh, 2006; Lee and Paull, 2007).

Telomere Repeat Binding Factor-2 (TRF-2) has a role in the activation of the ATM/p53 DNA damage response pathway (Karlseder et al. 1999). Studies by Raynaud *et al.* suggested that dysfunction within DNA repair pathways occur between pre-neoplasia and breast cancer and this is associated with telomere length abnormalities. ATM activation occurred in the pre-cancerous stage and not in invasive breast cancer onset, suggesting DNA damage pathways having a role in prevention of progression to malignancy. This was correlated with shorter

detectable telomere lengths when comparing pre-neoplastic lesions to normal tissue (P=0.0116) using *FISH*. There appeared to be a higher frequency of telomere length abnormalities in general in breast cancers than in pre-neoplastic lesions, with either very short or very long telomere length profiles (Raynaud et al., 2010). These bimodal distributions observed by *FISH* suggest the possibility of multiple mechanisms at play in the progression to malignancy.

4.2.4 DNA damage, BRCA-1 and telomere dynamics

Mutations in the breast cancer susceptibility genes 1 and 2 (BRCA1/BRCA2) are known so increase the risk of developing breast cancer considerably (60-85%) (Easton et al., 1993; Streuwing et al., 1996). Both BRCA 1 and BRCA 2 have been implicated as having roles in DNA repair mechanisms and thus play a role in maintaining the stability of the human genome (Cortez et al., 1999; Lou et al., 2003). BRCA1 is involved in the repair of DSB by homologous recombination by binding directly to DNA, and also via interactions with RAD51 (Scully et al., 1997; Cousineau et al., 2005). RAD51 is involved in the locating homologous regions and pairing up strands during the repair process [Reviewed by Fillipo et al., 2008]. BRCA2 also interacts with RAD51 to regulate both the intracellular localisation and DNA-binding ability of this protein (Chen et al., 1999).

BRCA1 and BRCA2 mutated tumours have been shown to result in gross chromosomal rearrangements (Venkitaraman, 2002; Gretarsdottir *et al*., 1998). Studies in mice have shown that disruption of BRCA2 results in chromosomal aberrations such as chromosomal breakage (Patel *et al*., 1998; Xu *et al*., 1999). Spectral karyotyping of BRCA 2 deficient cells has shown a multitude of chromosomal rearrangements such as translocations or deletions and evidence of NHEJ (Yu *et al*., 2000).

A study by Al-Wahiby S and Slijepcevic P, 2005, investigated BRCA1 deficient human and mouse cells for the presence of chromosomal aberrations suggestive of telomere dysfunction. A lymphoblastoid cell line, GM14090, (BRCA 1 heterozygote) showed an increase in detectable dicentric chromosomes. Molecular cytogenetic analysis revealed that these dicentric chromosomes arose as a consequence of end-to-end chromosome fusions due to the presence of telomere sequences at fusion points (Al-Wahiby S and Slijepcevic P, 2005).

4.2.5: BRCA-2 involvement in telomere maintenance

As mentioned previously, BRCA1 and BRCA2 are involved in the repair of DNA double stranded breaks (DSBs) via homologous recombination (HR) and thus mutations in either gene can drive genome instability. Studies have shown that cells with mutated copies of the BRCA2 gene display a variety of chromosomal abnormalities such as chromosome breakage, translocations and end fusions that are consistent with telomere dysfunction. BRCA2 has been implicated in telomere biology; Badie *et al.* showed that BRCA2 facilitates the loading of RAD51 onto telomeres as a mechanism of telomere replication and capping (Badie et al., 2010). BRCA2 deletion in mouse mammary tumours resulted in telomere shortening and the accumulation of phenotypes associated with telomere dysfunction such as fragmented telomere signals detected by *FISH*. Research has also shown an indication of telomere dysfunction in BRCA2 mutated human mammary tumours (Bodvarsdottir et al., 2011). Findings suggest that BRCA2 has a role in telomere capping and stabilisation due to the increase in chromosomal aberrations in the absence of functional BRCA2 genes (Bodvarsdottir et al., 2011). BRCA2 mutation carriers showed a higher frequency of chromosomal end fusions compared with breast tumours that did not have a BRCA2 mutation, suggesting BRCA2 loss was associated with telomere dysfunction. These data

implicate BRCA2 in having a role in telomere maintenance (Bodvarsdottir et al., 2011).

4.2.6: Telomerase activity and breast cancer

Due to its role in maintaining telomere length, telomerase expression has been linked with cellular immortality and cancer. Around 80% of malignant tumours overexpress telomerase (Kim et al., 1994; Shay et al., 1997). Correlations have been shown between telomerase expression and poor prognostic outcome in numerous types of cancer (Streuker et al., 2001; Artandi et al., 2009). A study by Hiyama *et al.* investigated the relative expression of telomerase in a range of different breast cancer subtypes. Findings from this indicated telomerase activity was detected in 68% of primary stage I breast cancers and 95% in the most advanced stage tumours. These observations point to a role for the activation of telomerase in later stages, which will result in stabilisation of telomere length and thus immortalisation of cells (Hiyama et al., 1996). A study by Clark *et al.* assessed telomerase activity in patients with node-positive breast cancer and this was linked to patient survival data (Clark et al., 1997). Findings from this study showed that increasing levels of telomerase expression were correlated with decreased disease free survival (P=0.041) and overall survival (P=0.009) (Clark et al., 1997). Mokbel *et al.* studied telomerase expression in different stages of breast cancer. In normal breast tissue, benign breast and ductal carcinoma *in situ* (DCIS) samples telomerase expression was undetected, however in infiltrating carcinomas 67% of samples showed expression of telomerase, indicating an association between telomerase activity and the progression to invasive malignancy (Mokbel et al., 1999). These studies indicate that telomere elongation by telomerase may be an important step in the progression to malignancy as a way of rendering cells immortal.

4.2.7 Project aims:

Studying telomere dynamics in breast cancer has revealed correlations between short telomeres and breast cancer progression. However direct evidence that telomeres within breast cancer cells can shorten to within the length ranges at which dysfunction occurs, is lacking. The aim of this study was to address this issue by using high-resolution single telomere length analysis (STELA) to provide a detailed description of telomere length distributions in breast cancer.

This work provided evidence of evidence of short telomeres within the length ranges at which telomere fusion can occur, evidence of clonal evolution and complete telomere loss.

4.3 Results:

DNA was extracted from human breast tissue by the Wales Cancer Bank. Unfortunately the

range of concentrations and their very dilute nature rendered them difficult to work with. Five of the 130 samples contained no detectable DNA (0ng/ul) when quantified using both Fluorometer and Nanodrop quantification. Concentrations of DNA ranged from 10ng/ul to around 75ng/ul and the average concentration detected was 31ng/ul. STELA is a single molecule technique that readily amplifies telomeres within a sample when input DNA concentrations are low (typically 10ng/ul). This meant that the DNA samples received were just adequate for successful STELA analysis, however additional reactions had to be undertaken in order to obtain a sufficient sample size of telomeric molecules. A full table of DNA concentrations is available in 'Chapter 2: Materials and Methods'.

4.3.1 STELA profiles

STELA is a powerful tool for the analysis of telomere length profiles *in vitro* and *in vivo* as it is a high resolution method for quantifying individual telomere lengths present in a sample; furthermore the distribution of these telomeres can be easily characterised in this way. Figure 4.1 shows a typical STELA blot generated for three patients comparing XpYp and 17p telomere length distributions. Figure 4.2 shows a summary of all telomere lengths detected at XpYp, 17p and 2p by STELA. As is visible in the scatter diagrams numerous patients had mean telomere lengths below 5kb. Mean telomere length varied by 8.39kb between the longest of 9.68kb and shortest 1.29kb at the XpYp telomere. The shortest telomere detected by TRF analysis, the gold standard for telomere length analysis, that has been recorded in the literature for breast carcinoma is 3.7 kb (Chin et al., 2001). Telomere lengths documented in the literature using TRF analysis tend not to be any lower than 4.0 kb, for example the shortest mean telomere length reported by Hiromi et al. is 5.2kb (measured using TRF analysis) (Schroder et al., 2001; Hiromi et al., 2011). Extensive analysis using STELA in CLL has defined a threshold length below which telomere fusion is detected (3.81kb for XpYp and 4.81kb for 17p-see red dotted line in figure 4.2). In the breast cancer cohort 57%

of tumours were within this length.

Figure 4.1: Comparing XpYp and 17p telomere lengths for patients 1-3 with IDC.

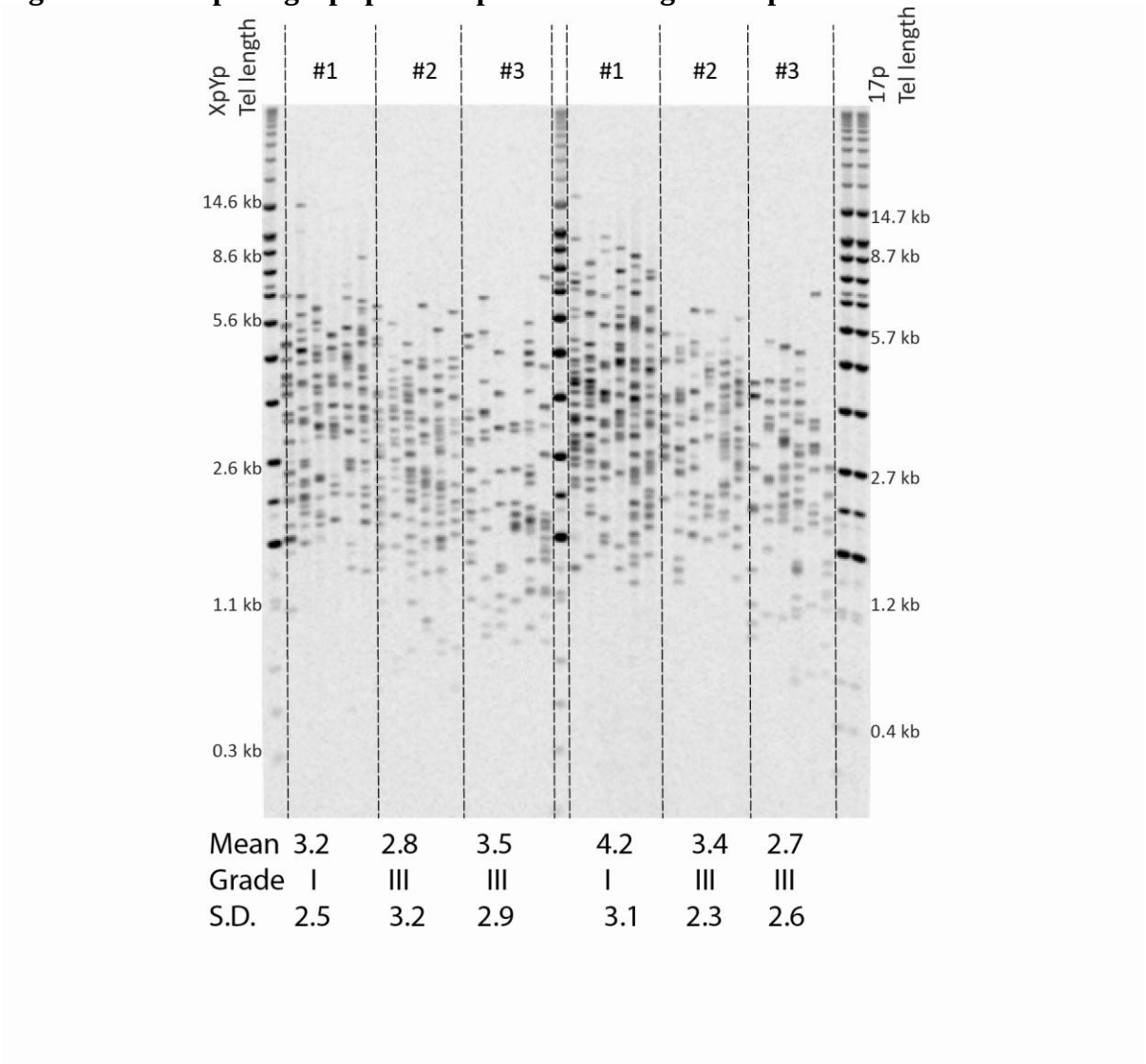
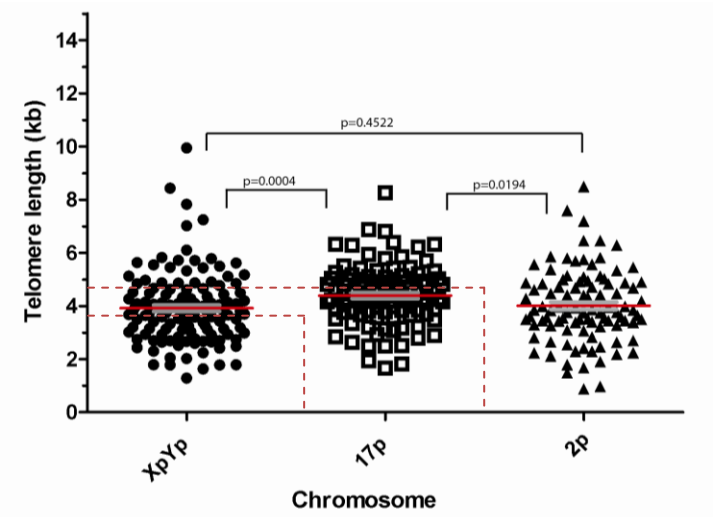


Figure 4.1: STELA for 3 patients with IDC. The left hand side of Southern blot shows telomere lengths for XpYp and the right hand side shows the same patient samples analysed for 17p. Mean, grade and standard deviation (S.D.) are shown for each patient below each telomere length profile.

Figure 4.2: Mean telomere length scatter plot



One-way analysis of variance	
P value	0.0206
P value summary	*
Are means signif. different? (P < 0.05)	Yes
Number of groups	3
F	3.928
R squared	0.02396
Bartlett's test for equal variances	
Bartlett's statistic (corrected)	3.897
P value	0.1425
P value summary	ns
Do the variances differ signif. (P < 0.05)	No

Figure 4.2: Figure shows a scatter plot of mean telomere lengths for each telomere analysed by STELA and the difference in means between those groups. CLL defined fusogenic lengths of 3.81kb for XpYp and 4.81kb for 17p is shown by red dotted line

4.4: Telomere lengths for three chromosome ends

Three chromosome ends were chosen as candidates for STELA analysis: XpYp, 17p and 2p. STELA was originally developed for the XpYp telomere due to its unique sequence and subsequently at the 17p telomere which also has a unique telomere sequence ideal for primer design; both these chromosome ends allow for a very robust STELA.

Chromosomal rearrangements such as deletions, amplifications and inversions are common at loci containing oncogenes, for example loss of 17p is frequent due to the presence of the tumour suppressor gene p53 (Cogen et al., 1990; White et al., 1996; Lansdorp et al., 1998).

Chromosomal abnormalities have been documented for chromosome 2p in the literature. Genomic rearrangements involving the losses and gains of chromosome 2p (specifically the 2p24 locus) have been shown to lead to MYCN amplification which confers poor prognosis in patients with neoroblastoma (Pandita et al., 2010). Genetic abnormalities have also been characterised extensively for the 2p chromosome locus in hereditary non-polyposis colorectal carcinoma (HNPCC) due to the presence of mismatch repair genes MSH2 and MSH6. Frequent deletions and rearrangements involving 2p have been observed in HNPCC (Wijnen et al., 1998a; Wagner et al., 2003; Wijnen et al., 2005) and the disruption of the mismatch repair genes found in this region is thought to be a contributor to the disease. MSH2 is involved in the recognition of DNA damaging lesions in the mismatch repair process (Fishel et al., 1994) and has been shown to interact with BRCA-1, with BRCA-1 acting as a downstream effector of DNA mismatch repair signalling pathways (Wang et al., 2000; Wang et al, 2001). The lifetime risk of ovarian cancers in patients with HNPCC is higher, however it remains unclear if this is true for breast cancer. Colorectal cancer mutation carriers have been shown to develop breast cancer and in a study it was observed that in patients with both HNPCC and breast cancer over 50% had defects in mismatch repair genes (Walsh et al., 2010).

Gains and losses involving chromosome 2p have also been observed in patients with breast cancer. In patients with a form of DCIS loss of heterozygosity was observed at multiple chromosome ends, one of which included 2p, using known polymorphic DNA markers in a PCR reaction (Moinfar et al., 1999). Array-CGH techniques have also been used to assess genetic alterations in an MCF7 and CL-9 breast cancer cell line (tamoxifen sensitive and tamoxifen resistant respectively) showing amplification of 2p regions (Achuthan et al., 2001).A 2p STELA assay has also been developed and used for the analysis of this telomere in cultured cells, but it has not been tested in a cohort of different individuals.

A comparison of the mean telomere lengths across all three chromosome ends showed no significant differences with One-way Anova analysis (statistically $p>1$) (figure 4.2).

Telomere length data for the 2p telomere appears to be more homogeneous (figure 4.2), and telomeres at XpYp have the largest degree of heterogeneity compared with 17p and 2p when analysing the mean data subjectively. The differences in standard deviation between the three telomeres are statistically significant (One way anova, $p=0.0002$) (Figure 4.4). Some of the STELA blots however, showed that fewer telomere molecules were detected at 17p and 2p than at XpYp. This is likely to be due to differences in the single molecule amplification efficiencies between the STELA assays at the three telomeres analysed. XY scatter plots revealed that the mean length of the XpYp and 17p telomeres was significantly correlated ($r^2=0.5$; $p<0.0001$; figure 4.3 (a)). The regression line crossed the Y-axis at 2.0 kb, indicating that the 17p telomere was on average longer than the XpYp telomere. However pairwise comparisons with 2p showed no correlation with 17p ($r^2=0.006$; $p=0.46$; figure 4.3 (b)) or XpYp ($r^2=0.0003$; $p=0.87$; figure 4.3(c)). This coupled with the low amplification efficiency meant that the data for 2p was considered unreliable and was not used in the subsequent analysis.

Figure 4.3: Comparison telomere length means between telomeres

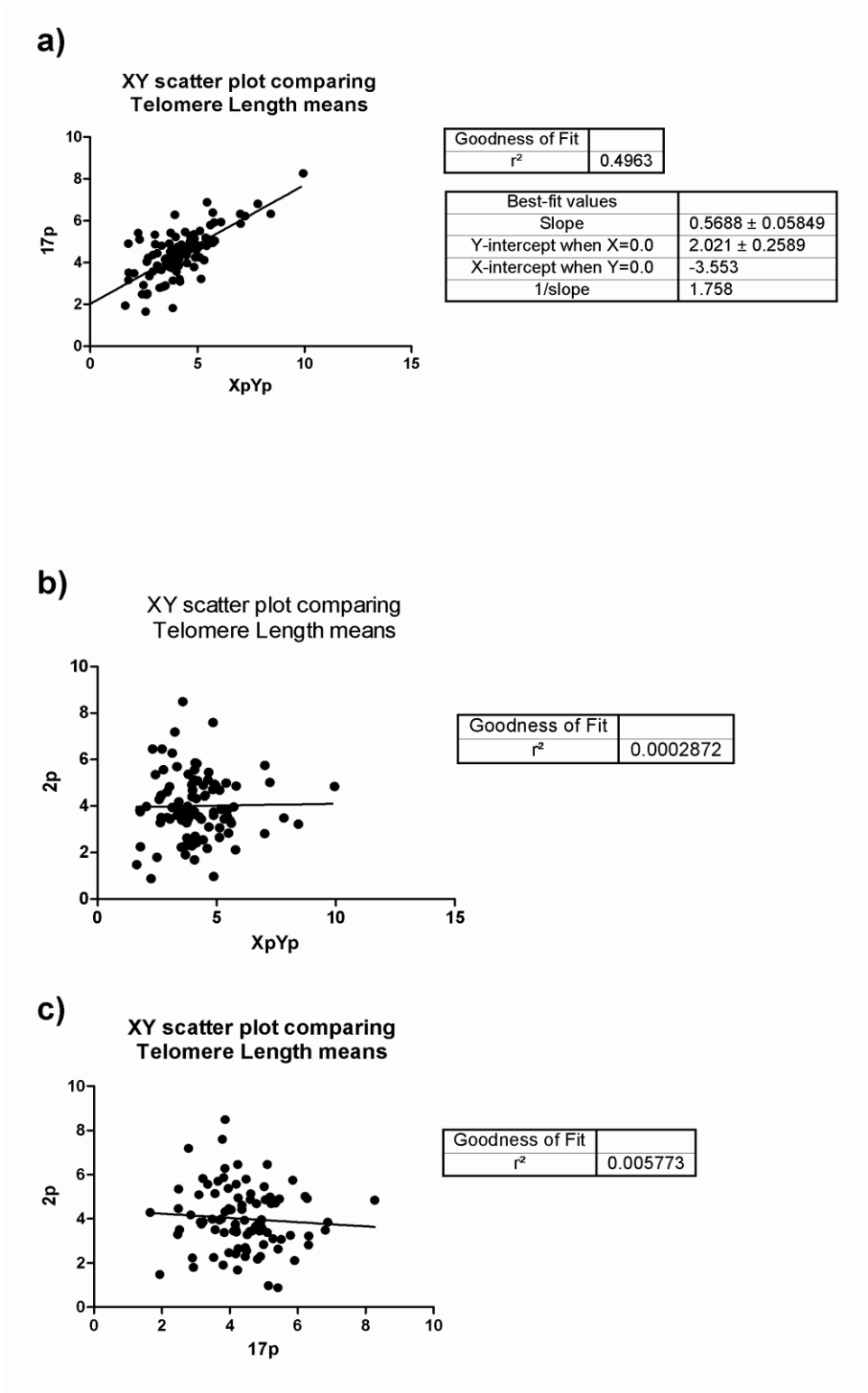
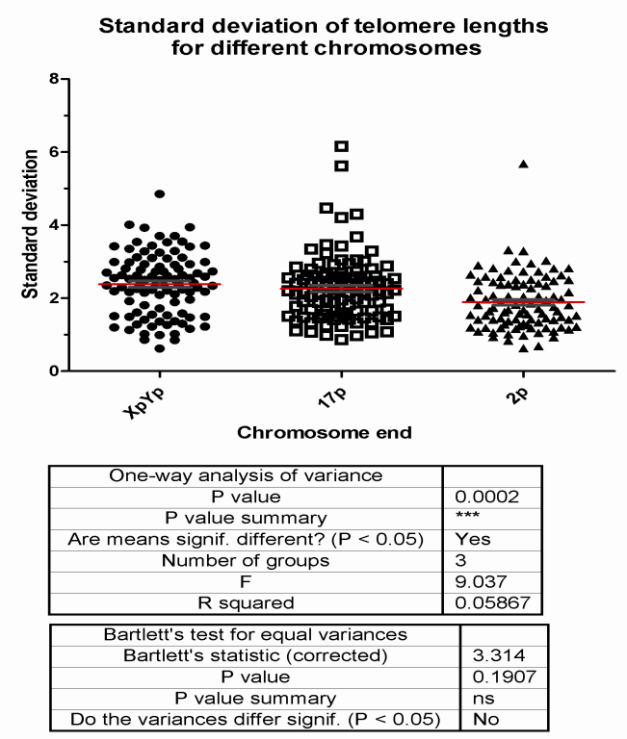


Figure 4.3: Figure shows a scatter plots comparing telomere lengths between telomeres to observe any similarity between telomeres. a) XpYp/17p (top) has the best r-squared value (0.49 ; $p<0.0001$). The regression line for this dataset crosses the y-axis at 2kb and thus 17p is as much as 2kb longer than XpYp. B) and c) show no correlation

Figure 4.4: Standard deviation for mean telomere lengths shown on a scatter plot



One-way analysis of variance	
P value	0.0002
P value summary	***
Are means signif. different? (P < 0.05)	Yes
Number of groups	3
F	9.037
R squared	0.05867

Bartlett's test for equal variances	
Bartlett's statistic (corrected)	3.314
P value	0.1907
P value summary	ns
Do the variances differ signif. (P < 0.05)	No

Figure 4.4: Figure shows a scatter plot of standard deviations for each mean telomere length, for each telomere analysed by STELA.

4.5 Telomere length clonality vs. heterogeneity

STELA analysis revealed several homogenous telomere-length distributions, with telomeres that appeared clustered together in a 'clonal' pattern (Figure 4.5b), whereby the standard deviations were low and the difference between longest and shortest telomere for an individual was around 1kb or less. Clonal telomere distributions have been observed previously in culture-models (Baird 2003; Britt Compton 2006) in CLL using STELA (Lin et al., 2010). Short clonal telomere distributions similar to fibroblast populations derived from a single cell were hypothesised to be consistent with clonal B-cell growth (Lin et al., 2010). Binet's staging is the gold standard staging criteria used throughout Europe that classifies CLL patients into three groups (A, B and C) depending on lymphoid involvement and the

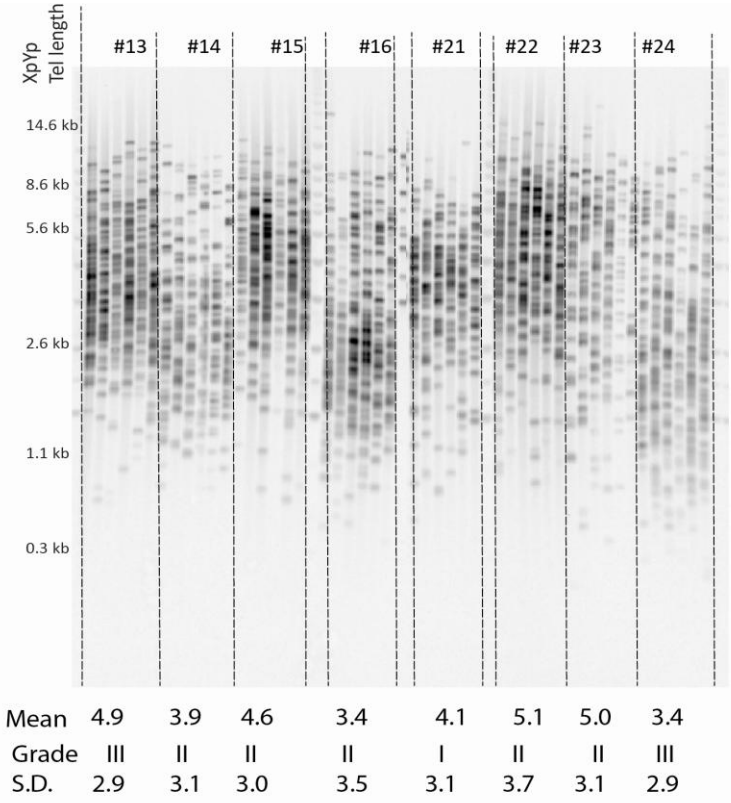
presence of anemia and/or thrombocytopenia. Patients with the most severe prognosis according to Binet's staging (stage C) had the shortest telomere distributions and the most clonal distributions compared to other stages (A and B), indicating extensive cellular proliferation alongside hallmarks of genomic instability such as LOH consistent with the accumulation of instabilities giving rise to the clonal evolution of the B-cell population.

In contrast other tumours displayed more heterogeneous telomere-length distributions with much higher standard deviations (figure 4.5a). Genetic heterogeneity can arise in cancer cell populations for numerous reasons and these telomere length distributions may be consistent with oligo-clonal growth. Microsatellite instability due to defective mismatch repair mechanisms can give rise to genomic heterogeneity, and may generate telomere length heterogeneity (Pickett et al. 2004) Other changes such as single nucleotides polymorphisms (SNPS) and copy number alterations (CNAs) create heterogeneity [Reviewed by Miller et al., 2003] and are frequently observed in tumour via array-CGH technology that compares tumour DNA with normal DNA via comparative microarray hybridisation with known DNA probes (with a known genomic position eg. cDNA libraries) (Kallioniemi et al., 1992).

A further explanation for the presence of heterogeneous telomere length distributions is the cell purity of the tissue samples analysed. The presence of stromal cells or infiltrating lymphocytes that are likely to contain longer telomere length will influence the telomere distribution from the tumour samples, this is likely resulting in an over estimation of telomere length. STELA analysis of the breast cancer cohort thus revealed an array of different telomere length profiles, some of which are summarised by use of scatter diagrams in figure 4.6.

Figure 4.5 Clonal vs heterogeneous telomere length distributions

a)



b)

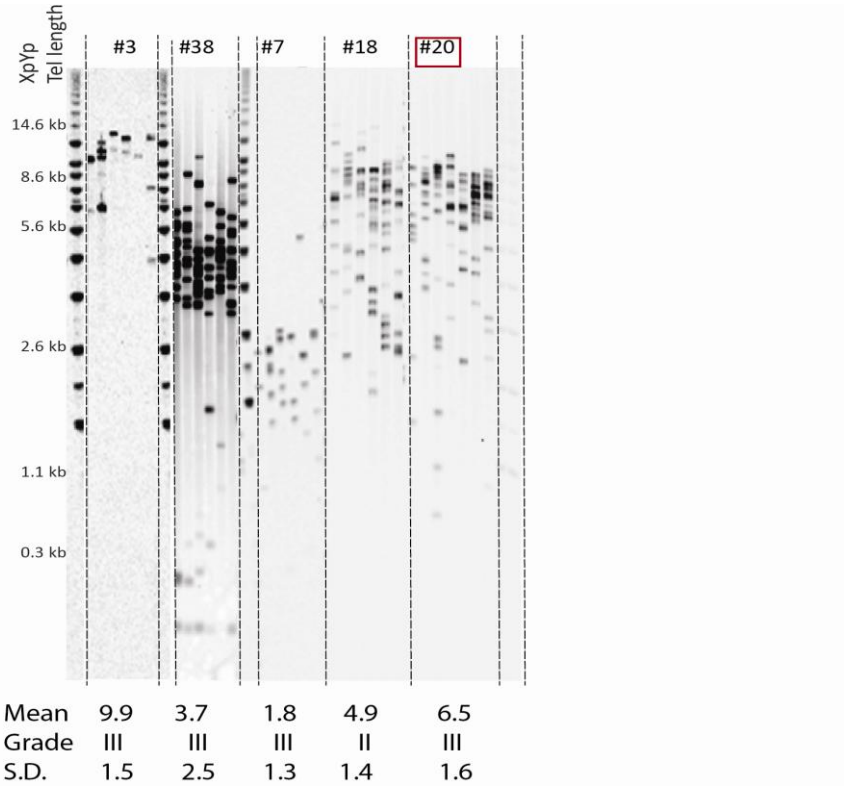


Figure 4.5: Figure showing the contrast between heterogeneous telomere length profiles (top-**a**) and homogeneous 'clonal' profiles (bottom-**b**). Telomere mean was calculated for each patient with the histological grade and standard deviation (S.D.). The red box around #20 depicts a patient that has died from invasive ductal carcinoma.

Figure 4.6: Variability in Telomere length profiles

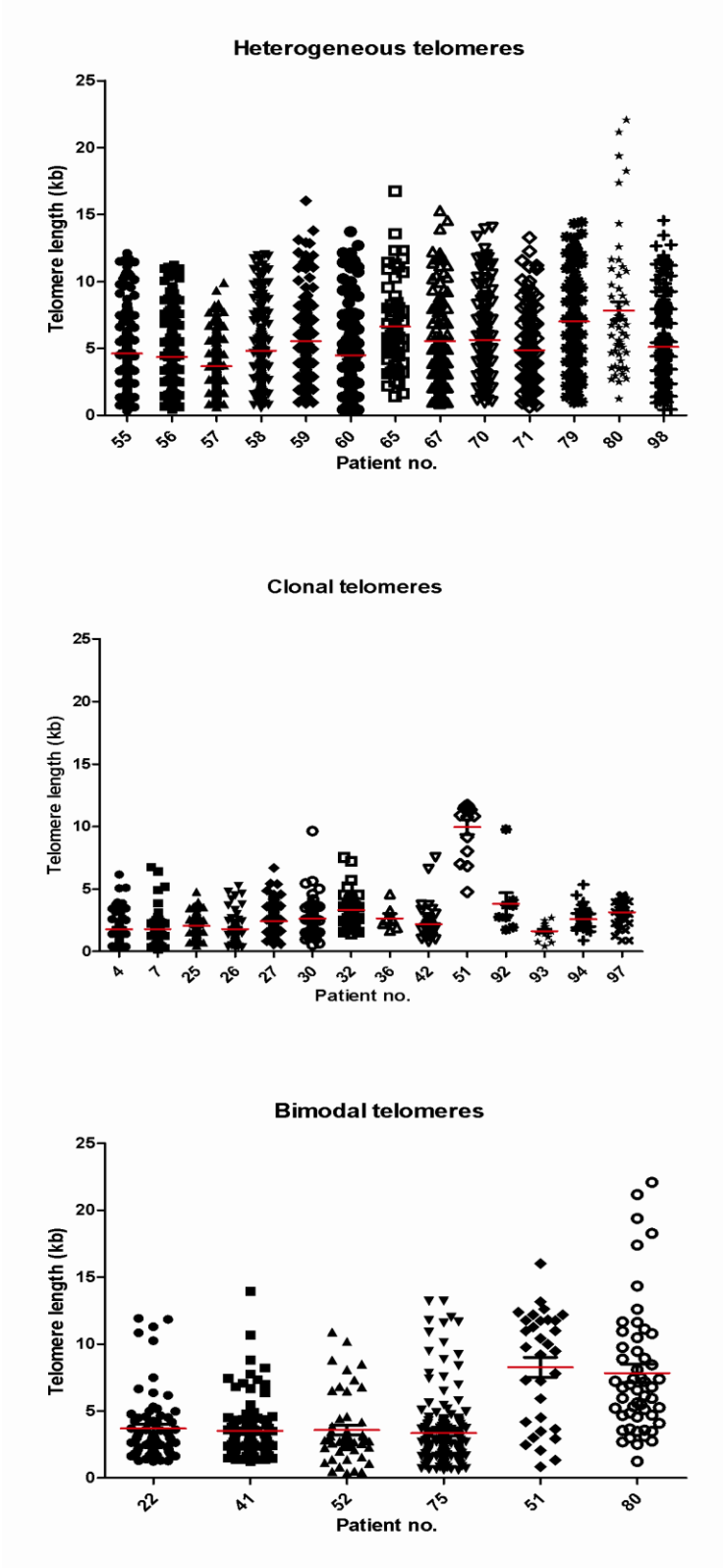


Figure 4.6: A summary of the most heterogeneous (top), homogeneous (middle) and bimodal (bottom) telomere lengths from the patient cohort of 130 are illustrated by the use of scatter plots.

4.6 Bimodal distributions

One striking observation was that several tumour samples exhibited clear bimodal telomere length distributions (figure 4.7). The underlying cause of these distributions is not clear. One possibility is intra tumour heterogeneity. Alternatively contaminating somatic tissue with long telomeres could influence the resulting telomere length distributions. As seen previously when analysing STELA blots, numerous patients appeared to have bimodal telomere length distributions, two clear clusters of populations of telomere lengths with distinctly different means (see figure 4.7 for an example of this). Bimodality has been observed in breast tumours in relation to ER expression with some tumours displaying both ER+ and ER- subtypes (Schnitt et al., 2006). However there is controversy whether tumours of this type are actually bimodal for gene expression or whether inaccuracies with immunohistochemistry analysis are providing false results of just artefacts making tumours appear bimodal (Rimm et al., 2007). Whether the telomere distributions are representative of bimodal tumours or not, it is difficult to deduce the mean telomere length accurately, longer or shorter sub-clusters of telomeres could skew the mean. It was hypothesised that non-breast cancer tumour DNA could have an effect on the telomeres detected, this is because removing tumour DNA will incorporate other somatic cells and is not specific to just breast tumour DNA. Infiltrating lymphocytes within a sample express telomerase, which will elongate and thus clusters of longer telomeres have the potential to skew the results.

Figure 4.7: Bimodal Telomere length distributions at XpYp for IDC patients

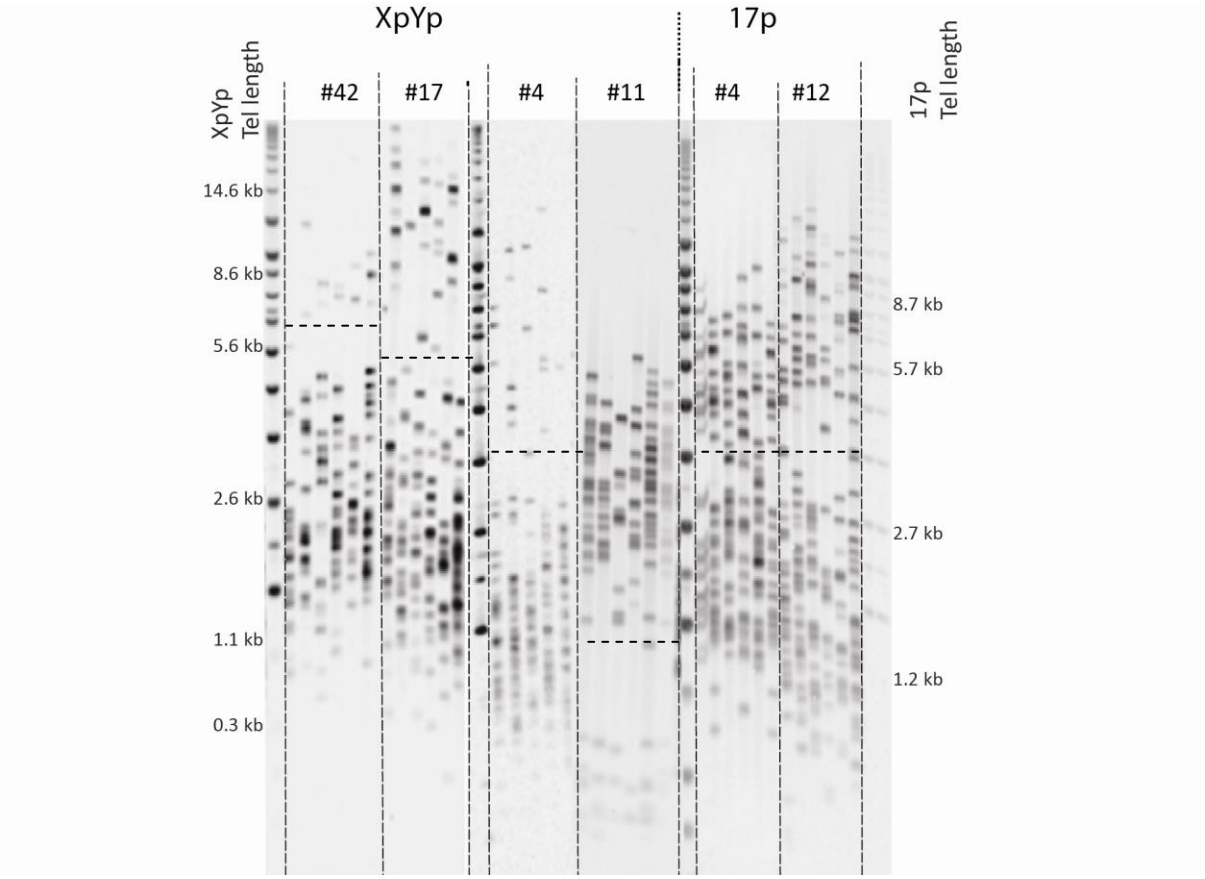


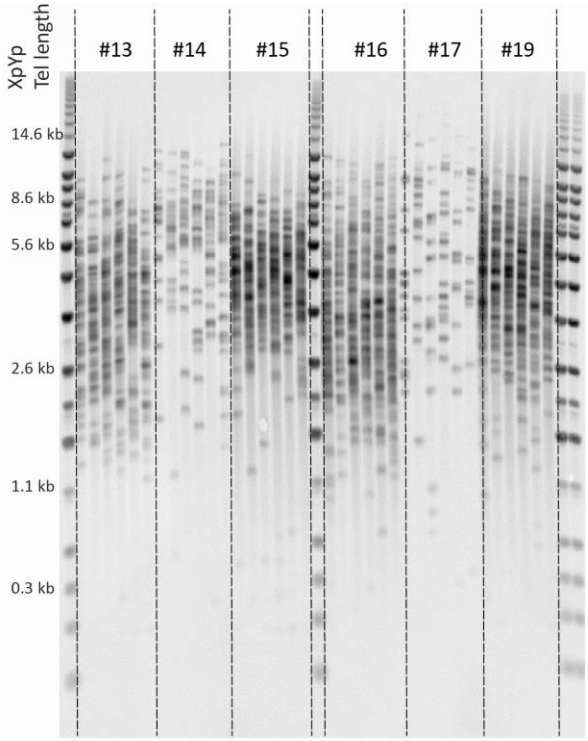
Figure 4.7: Figure shows STELA blot with bimodality. The dotted horizontal lines indicate the divide between clusters of long and short telomere length distributions

4.7 Longer mean telomere length distributions at XpYp and 17p

STELA blots were compared against clinical features such as grade. There were a profusion of different profiles of telomeres for each subgroup, for example some patients with the highest grade (grade III) and hence a more severe prognosis actually had a longer mean telomere length distribution (patient #14, figure 4.8) than those with a less pronounced

disease.

Figure 4.8: Long telomere distributions at XpYp for patients with IDC



Mean	4.8	8.4	7.0	5.9	6.3	7.2
Grade	I	III	II	II	II	II
S.D	3.1	3.2	3.9	3.0	3.5	3.4

Figure 4.8: Figure shows a STELA blot with the longest patient telomere distributions. Below this southern blot is a table with the telomere mean calculated for each patient with the histological grade.

4.8: Loss of heterozygosity (LOH)

Detection of loss of heterozygosity (LOH) can be a way to find putative tumour suppressor genes involved in the progression to malignancy in cancer. LOH is when one or both arms of certain chromosomes are deleted due to the loss of a normal functional allele at a heterozygous locus (Callahan et al., 1993; Miller et al., 2003). This can occur if there is a

tumour suppressor gene that is mutated to facilitate tumour progression, and thus the mutation/loss of function of this gene results in subsequent loss of alleles. Pooled analysis of LOH occurrence in breast cancer has revealed favoured loss at multiple chromosome arms, including 17p (Miller et al., 2003). This is hypothesised to be a common locus for LOH in many types of cancer due to the presence of the p53 gene at 17p13.1 (McBride et al., 1986).

STELA blot analysis has revealed varying degrees of LOH for XpYp and 17p by quantification of the number of amplifiable molecules at each end. One patient exhibited loss of both telomeric alleles at 17p (figure 4.9a). Several other patients displayed a telomere loss of between 30-60% for both XpYp and 17p telomeres. Patients with above 50% LOH at 17p were tabulated (table 4.1) and the differences in telomere molecules is highlighted by use of a scatter plot (Figure 4.9b)

Figure 4.9.a Example of complete LOH of 17p telomere

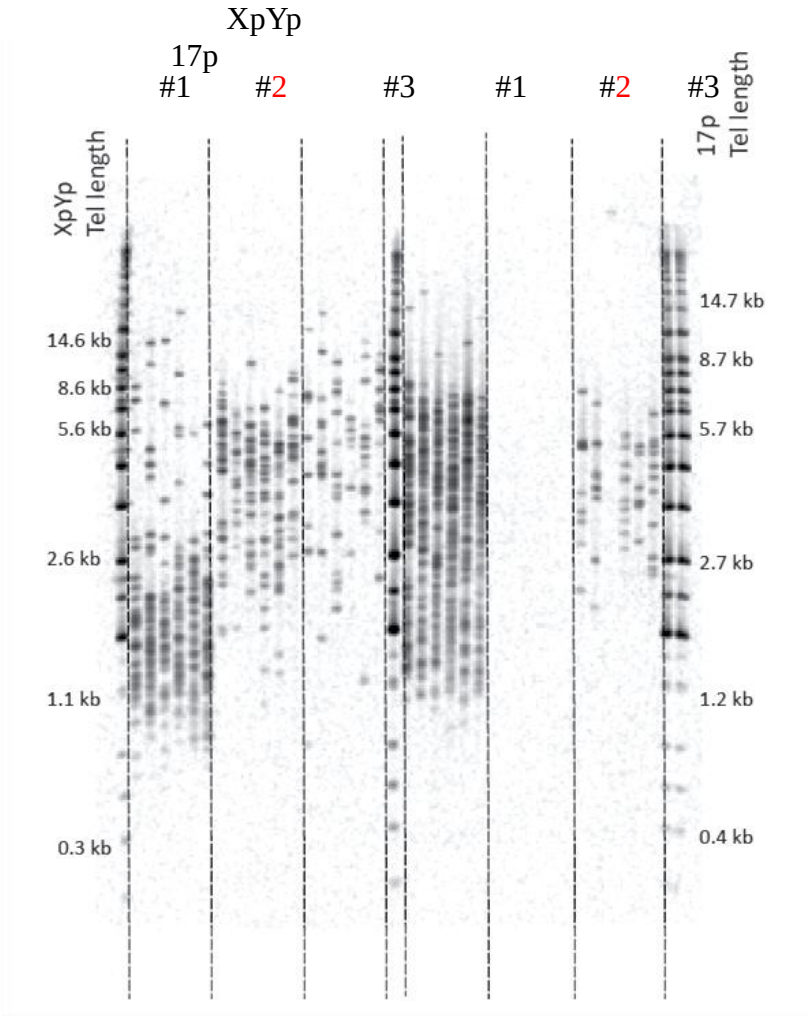


Figure 4.9a: Figure shows STELA blot with an example of loss of heterozygosity (LOH). Patient #2 (shown in red) has a complete loss of the 17p allele visible by southern blot analysis (STELA).

Figure 4.9b: LOH comparison for XpYp and 17p

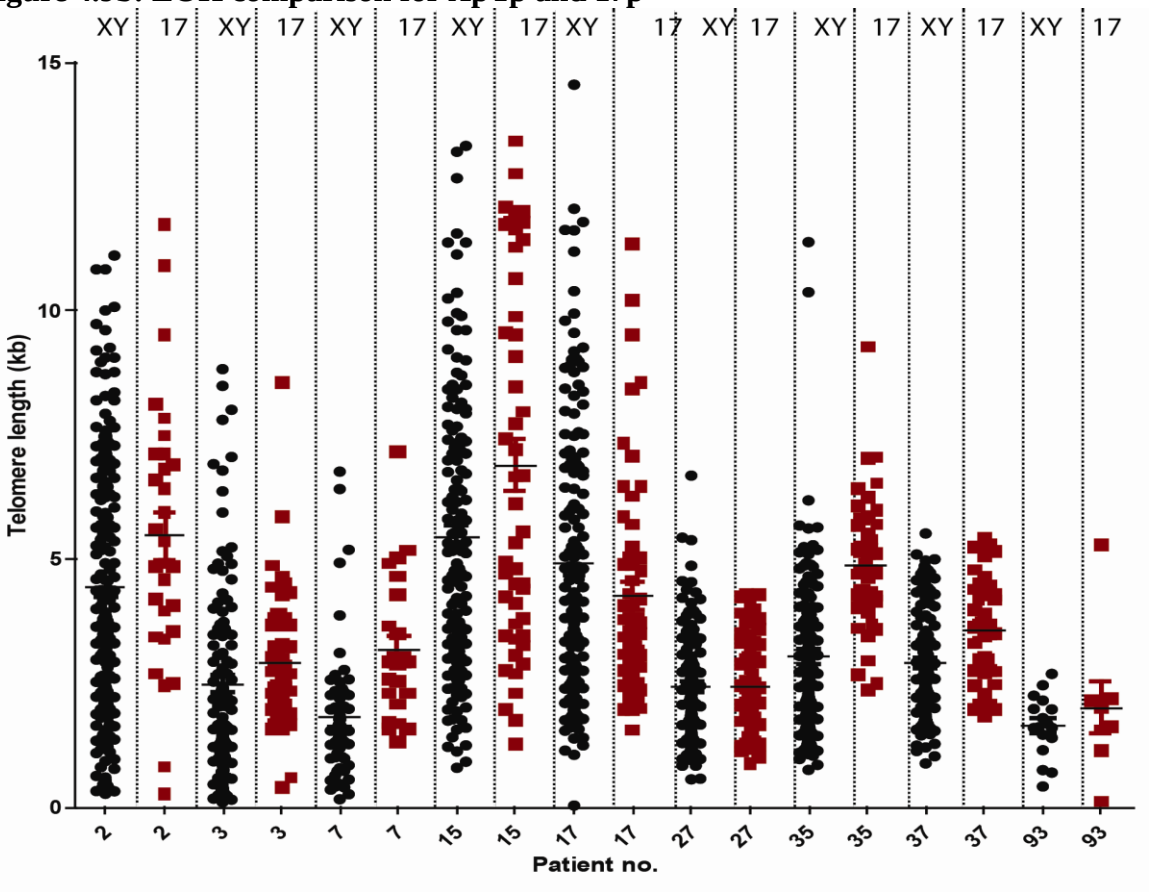


Figure 4.9b: Red squares indicate telomeres detected at 17p and black circles indicate telomeres detecxted at XpYp.

Table 4.1: Patient LOH

Patient Number	Molecules at XpYp	Molecules at 17p	LOH detected
2	204	30	85%
3	120	46	62%
7	64	23	64%
16	138	49	64.5%
18	158	59	63%
29	117	67	62%
36	131	43	67%
38	104	39	62.5%
94	17	8	53%

Table 4.1: A summary of the patients with high percentages of LOH for 17p when compared with XpYp.

4.9: Discussion:

4.9.1 Telomere length analysis in Invasive Ductal Carcinoma

Applying STELA to a cohort of 130 patients with IDC has allowed the high resolution analysis of individual telomere lengths, revealing a spectrum of telomere length distributions. There is no clear pattern of telomere lengths found in IDC. Some telomere length profiles had a low standard deviation that was consistent with clonal growth, similar to that observed in single-clones in culture or in CLL patients; others displayed heterogeneous telomere–length profile; whilst others appear to have two distinct groups of telomeres with distinct mean lengths (bimodal populations). Short telomere length has been correlated with prognostic severity in numerous cancers including breast cancer. The utilisation of this high resolution single molecule technique for measuring telomeres has provided evidence of extremely short telomeres in this malignancy.

4.9.2 Telomere length differences between XpYp and 17p

Mean telomere lengths at 17p were longer than XpYp when comparing using a linear regression analysis ($P<0.0001$; See figure 4.2). Is has been documented previously that the 17p telomere has a propensity to be shorter than other chromosome ends (Martens et al., 1998, Lansdorp et al., 1998). It has been postulated this it may be related to the presence of the oncogene p53 on the 17p chromosome arm. Studies have suggested that the reason for this difference in length is that shorter telomeres act as a molecular ‘clock’ for replicative senescence (Martens et al., 1998). However other studies have shown that 17p is not always the shortest telomere, and that some telomeres are randomly shorter (Baird et al., 2006) suggesting telomere length is governed at the zygote stage and variability could be due to

difference in the maternal and paternal germlines. 17p has been postulated to be shorter in many cell types and enabling a protective mechanism due to the presence of p53; if 17p is set shorter than the rest of the telomeres then the cell can senesce/exit the cell cycle more rapidly preventing mutation at p53 (Baird et al., 2006). Overall the 17p data here conflicts with the idea that 17p is shorter in cancer (Lansdorp, 1998). An alternative explanation for this observation may be that presence of Telomere Variant Repeats (TVRs) at 17p may make the telomeres appear to be longer, i.e. there might be more TVRs on average at 17p than at XpYp. A way of determining this in future would be to perform a TVR-PCR and deduct this from the STELA lengths. A recent study in CML (Chronic Myelogenous Leukemia) reported longer 17p telomere lengths in patients with CML when compared with healthy patient telomere lengths using Q-*FISH* (Samassekou et al., 2011).

4.9.3 The use of STELA in a large cohort study

STELA as an assay for measuring telomere lengths has allowed the detection of extremely short telomeres of less than 1kb accurately. However, the application of this technique in a large cohort study is time consuming as it is a multistep process that requires a long range PCR, a southern blot and then an overnight hybridisation with radiolabelled probes. Furthermore at present only a few telomeres can be analysed using the STELA technique. Groups of telomeres share large degrees of subtelomeric homology and this makes the creation of unique oligonucleotide primers for use in the long range PCR difficult and inaccurate for many telomere families. However, STELA is the highest resolution assay for detecting the shortest telomeres within a sample and thus the development of a high throughput STELA method would be beneficial for larger-scale epidemiological studies. Other methods of measuring telomere length are generally hybridisation based and cannot

detect telomeres below lower thresholds of TTAGGG repeats. High throughput Q-PCR based techniques only provide average telomere lengths per sample and individual telomeres cannot be quantified (Cawthon, 2002).

4.9.4 Telomere length analysis in breast cancer:

There were numerous different telomere length profiles: some telomeres appeared to be present in two distinct clusters at different molecular weights (bimodal). This could be due to allelic differences in length that have been seen in other cell types such as senescent human fibroblasts (Baird et al., 2003). However this appears unlikely as the presence of telomerase activity can homogenise telomere length distributions and thus remove allelic length differences (Britt-Compton et al. 2009). Bimodality could also arise due to intra tumour heterogeneity (as mentioned earlier). We considered that these bimodal distributions of telomeres could be due to the presence of other cell types contaminating the sample, such as telomerase expressing lymphocytes and stromal cells from the surrounding tissue. Due to the limited nature of the samples received and it was difficult to assess intra tumour heterogeneity as a means of bimodality, however future work with tumour tissue instead of DNA extraction samples would allow more detailed analysis of each breast tumour. The samples obtained from the Wales Cancer Bank were DNA extractions from breast tissue and had therefore not undergone any cell purification to isolate carcinoma cells prior to DNA extraction and this appeared to be the most likely explanation for the bimodality.

A number of telomere lengths detected by STELA were extremely short compared to what has been observed previously using TRF analysis (1.29kb here compared with lengths of 5kb

and above with TRF (Schroder et al., 2001; Hiromi et al., 2011). This suggests that the telomeres have eroded extensively, this is particularly apparent when one considers that non-functional telomere variant repeat regions that encompass 0-2 kb of the proximal regions of the telomere repeat array. Recent studies in CLL using the same STELA assay have shown telomeres that have eroded to a length reminiscent of cells in crisis *in vitro* (Lin et al., 2010). Telomere erosion may be a result of hyper-proliferation of cancer cells (Weiner et al., 2008). Mutations that target cell surface receptors are may drive proliferation in breast cancer cells, such as mutations in the ERBB2 gene causing overexpression of HER2 (Yu et al., 2007) and also mutations causing overexpression of ER as oestrogen is required for proliferation in normal breast tissue (Russo et al., 1999; Haslam and Woodward, 2003), thus overexpression will lead to uncontrolled proliferation and thus rapid telomere shortening. It has been shown that more aggressive subtypes such as HER2 positive have shorter telomeres (Heaphy et al., 2010). In CLL the most eroded telomeres were the ones undergoing the highest frequency of telomere-telomere fusions, a hallmark of telomere dysfunction (Lin et al., 2010). Thus eroded telomeres in the IDC samples described here could be eroded to the point at which they are dysfunctional, and telomere fusion could be causing large scale genomic instability thus driving the malignancy and not just a result of the proliferative capacity of cancer cells.

The telomere length data obtained from breast cancer samples has been consistent with previous reports using STELA in other malignancies. The shortest mean telomere length that was detected was 1.29kb (XpYp) which is the shortest telomere length detected for this disease. Single telomere length distributions (STELA) at XpYp and 17p have been previously examined in patients with CLL (Lin *et al*, 2010). These data showed that telomere length was shorter as a function of desease staging and a reduced heterogeneity of the length

distributions was observed, that was consistent with clonal B-cell growth. Functional telomere length was thought to be overestimated due to presence of TVRs, thus TVR-PCR was carried out, showing even more extensive telomere erosion during progression to malignancy. Array-CGH was also used for CLL samples and was used to test whether individuals showing telomere dysfunction showed genomic instability. There was clear evidence of large scale genomic instability that included NRTs in sample with telomere dysfunction. The application of TVR-PCR and array-CGH was not possible for the IDC samples due to the requirement for at least 100ng/ul of input DNA and nature of the samples received for experiments was insufficient. DNA aliquots received from the Wales Cancer Bank typically had as little as 10ng/ul of breast tumour DNA and a maximum of 60ng/ul in a volume of between 50 and 150ul. Development of the TVR-PCR assay as a nested PCR using more dilute input DNA was tested but unsuccessful. Array-CGH also requires a large concentration of starting DNA and thus could not be used for the DNA material provided to us. Nonetheless, telomere lengths detectable for IDC by STELA has allowed the detection of short telomeres and considering that part of this length would consist of telomere variant repeat regions, the lengths detected by STELA are extremely short.

4.9.4.1 Telomerase activity

Short telomeres found in this invasive form of breast cancer may be seen as contradiction of what would be expected when taking into account the high telomerase activity observed in the majority of human cancers. One model proposed for this would be that short telomeres are maintained by telomerase but not elongated as has been seen in HEK293 cells (transformed with adenoviral DNA) that express telomerase but maintain telomere length distributions that extend to within the length range telomere fusion events have been detected (Capper et al., 2007; Letsolo et al., 2010). In a small cohort study (n=11) with T-cell

prolymphocytic leukaemia (T-PLL) short telomeres were shown to have a high expression of telomerase (Roth et al., 2007) and a telomerase inhibitor (BIBR1532) led to selective death of the malignant T-PLL cells in vitro (Roth et al, 2007). Taken together it has been hypothesised that if a telomere is critically short, it would be prevented from shortening further by high expression of telomerase and this is why telomerase inhibitors will selectively kill the more aggressive cells because telomeres are then allowed to shorten to lengths recognised as critical by the cell (Lin et al., 2010, Jones et al., 2012). Telomerase activity in breast cancer patients from this cohort would thus provide more detail to elucidate potential mechanisms driving the progression to malignancy.

4.9.5 Loss of heterozygosity

Interesting observations can be made about the distributions of telomeres using STELA techniques, such as heterogeneity of samples, bimodality, clonality and loss of heterozygosity. One STELA profile in particular showed the complete loss of both alleles at 17p. If this extends to the p53 on 17p then the loss of the telomere could be related to the loss of p53. LOH of the short arm of chromosome 17 has also been observed in CLL clonal B-cells and has been linked with the loss of p53 (Lin et al., 2010). Therefore the loss of p53 could be a contributing factor to the progression to malignancy in IDC allowing neoplastic cells to escape senescence. However the 17p telomere is polymorphic in the human population (Britt Compton et al., 2006) and this may account for the lack of 17p telomere in these breast cancer samples. Unfortunately the absence of patient matched normal tissue samples meant that it was not possible to formally verify if the absence of the 17p telomere was a function of LOH events or a naturally occur polymorphism.

4.9.6 Telomere elongation

It's interesting to note that there was such a variation in the average telomere length detectable by STELA, all patients have the same invasive form of ductal carcinoma and there are still variations within groups of histological grade. One such observation was the presence of long telomere distributions, even in patients that had the worst prognostic outcome. This could be representative of telomere elongation events by the action of telomerase. 88% of ductal carcinomas have been shown to express telomerase (via the TRAP assay). Telomerase expression assays have also revealed a progressive increase in telomerase expression levels correlating with the severity of histopathological grade in breast cancer with 14% of telomerase positive samples in benign breast diseases, 92% in carcinoma *in situ* and 94% in invasive breast carcinomas (Yashima *et al.,* 1998). It has been proposed that telomere elongation can occur via a tankyrase-mediated ADP ribosylation of TRF1 that opens up the t-loop and thus facilitating the access to telomerase to add telomere repeats *de novo* (Lange *et al.,* 2000). Tankyrase expression levels have been shown to be higher in breast cancer tissue compared with normal tissue (Gelmini *et al.,* 2004). Interestingly this research has displayed that tankyrase overexpression is highly correlated with progesterone receptor (PR) positive breast tumours but not estrogen receptor (ER) positive breast tumours. No correlationcan be concluded from this project dataset, the longest two telomere length profiles were receptor negative for ER and PR.

The Alternative Lengthing of Telomeres (ALT) mechanism is relatively common in sarcomas for lengthening telomeres independently of telomerase activity. The ALT phenotype is, however, thought to be rare in carcinomas. ALT can be distinguished by the presence of promylocytic leukemia (PML) protein nuclear bodies (APBs) that contain large amounts of

extra-chromosomal telomeric DNA among other replication proteins. In a study by Subhawong *et al.* a subset of cancers with varying receptor status including HER2 overexpression were subjected to *FISH* combined with PML protein immunofluorescence. The ALT phenotype was identified in 3 out of 21 HER-2 positive cases compared and showed a high proliferative rate. None of the other subtypes of breast cancer showed the ALT phenotype and this was statistically significant. Two out of the three cases with the ALT phenotype had rapid tumour progression resulting in death (Subhawong et al, 2010). This suggests ALT may adversely affect survival outcomes. The mechanism underlying ALT is unknown, however it is hypothesised that a recombination pathway is involved due to the identification of proteins that are involved in homologous recombination in telomeres that are undergoing ALT. Phenotypically ALT telomeres have been shown to be long and heterogeneous (Subhawong et al, 2010). Some of the STELA profiles seen in this dataset of invasive ductal carcinoma are long and heterogeneous, for example patient #16 in figure 4.8 however it is not possible to determine if this is an example of ALT or heterogeneity generated by an independent mechanism. 29 out of 130 patients in this IDC cohort tested positive for HER2, of these 13 patients had standard deviations of above 2.5 and would be considered fairly heterogeneous (data shown in table in Appendix 2; Table B). One patient in particular had a standard deviation of 3.94 and a mean telomere length of 7.01kb; this distribution is fairly long and heterogeneous (patient #67 figure 4.3). This correlates with the previous evidence that HER2 expressing cells show indications of ALT mechanisms (Subhawong et al, 2010). Therefore future experiments could involve the testing of Her2 positive patient samples for the presence of ALT markers and this could elucidate telomere maintenance mechanisms at play to explain the heterogeneous distributions.

4.10 Conclusions and future work

Application of STELA for telomere length analysis has allowed the high resolution characterisation of telomere dynamics in invasive ductal breast carcinoma (IDC) for XpYp and 17p. The telomere length profiles differed markedly, with clonal, heterogeneous and bimodal distributions. More analysis of these telomere distributions is required to elucidate any mechanisms causing the differences in telomere lengths between patient samples for example: testing of patient samples for the presence of ALT markers in heterogeneous populations; allele specific STELA to examine the bimodality; and purification of the tumour tissue prior to DNA extraction to ensure tumour cells only within a sample.

The telomere lengths analysed for this cohort are the shortest described in a breast cancer cohort and are consistent with telomere lengths detected in CLL using the same technique. There is thus direct evidence for the presence of short telomeres in an invasive form of breast cancer which could be contributing to the progression of this disease. Future work with larger cohorts and more clinical stages could reveal the prognostic power of STELA for breast cancer and to elucidate whether short telomeres contribute to the progression of the disease.

It is pertinent to note that the telomere length analysis carried out by STELA only detects telomere lengths at XpYp and 17p. This is not representative of all telomeres within a sample, but only at those chromosome ends. Extending the STELA assay to encompass all telomere ends in the future would enable a more accurate estimation of mean telomere length for each patient. However the resolution of this assay with its ability to detect individual telomeres within a sample renders it extremely accurate for those chromosome ends.

Chapter 5: Clinical data analysis of commonly used biomarkers for breast cancer correlated with telomere length data acquired by STELA

5.1: Abstract

Clinical data associated with the cohort of IDC patients was received from the Wales Cancer Bank from patient follow up over a period of five years. This data included receptor status of all patients alongside survival outcome.

Telomere lengths for all patients were analysed using STELA analysis and the means were inputted into a spreadsheet containing all patient clinical information collected. This allowed the comparison of telomere length with clinical features, and these findings could then be statistically related to survival and thus prognostic value of telomere length in patients with different cancer subtypes/phenotypes.

Telomere lengths for different subtypes of breast cancer were analysed using Kaplan-Meier survival curves. Stratification of patients based on the threshold telomere length at which fusion occurs revealed a remarkable level of prognostic power with hazard ratios for overall survival that were orders of magnitude greater than that observed in the commonly used markers for prognosis in this disease.

5.1.2 Clinical background: Breast Cancer subtypes

Breast cancer can be classified by receptor status. ER+ (oestrogen receptor positive) and PR+ (progesterone receptor positive) breast cancer cells depend on presence of each hormone for their growth, respectively, thus drugs targeting oestrogen and progesterone receptors, such as oestrogen receptor antagonist Tamoxifen, can be used to block the hormone binding result in a block in cell proliferation and therefore a better prognosis.

In about 20% to 25% of breast cancers, the cancer cells overexpress a gene called Human Epidermal Growth Factor Receptor 2 or HER2 that is encoded by ERBB2, a proto oncogene located on the 17q chromosome. The ERBB2 gene is overexpressed in around 30% of breast cancers (Yu et al., 2007). Patients with a mutated ERBB2 gene have been shown to have a worse prognosis than ER+ patients; however monoclonal antibody treatment with Trastuzumab or Pertuzumab can help to improve this [Reveiwed by Weigelt et al., 2005]. HER2 is sometimes referred to as ‘neu’ because of it was characterised from a rodent neural tumour cell line (Coussens et al., 1985). Breast cancers with overexpression of HER2 tend to be much more aggressive and fast-growing (Tan and Yu, 2007). HER2+ breast cancer patients generally have been shown to have a worse prognosis than ER+ patients, however monoclonal antibody treatment can help to improve this [Reviewed by Weiglet et al., 2005 and Dunnwald et al., 2007].

Cells lacking ER/PR expression and also HER2 overexpression are referred to as basal-like or triple negative. The majority of breast cancers associated with the breast cancer gene known as BRCA1 are triple negative (Anders et al., 2008). About 75% of all breast cancers are ER positive and roughly 65% of these are also PR positive and thus grow in response to progesterone (Statistics retrieved from American Cancer Society FDA 2012 URL:Breastcancer.org; Dunnwald et al., 2007).

As mentioned previously, correlations have been shown between short telomere length and the more aggressive breast cancer subtypes, such as HER2 and triple negative (Heaphy et al., 2010).

The heterogeneity of breast cancer means that there is a high diversity in the tumour types found in patients. This makes it difficult to treat patients and predict their clinical outcome.

5.1.3 Breast cancer metastasis markers: requirements for better prognosis

The main cause of death from breast cancer in the Western world is from metastasis of malignant cells, not the primary tumour alone due to the difficulty in treating metastatic cells once they have entered the circulatory system and travelled to distant organs/tissues. It is estimated that 20% to 50% patients will develop metastatic breast cancer (Yu et al., 2009) and only 26% of these will survive past 5 years. Cytotoxic therapies, such as chemotherapy, are still the first port of call for treating metastatic breast cancer, or adjuvant therapies for use post-surgery to prevent local recurrence. Several different drugs are used as a chemotherapeutics for IDC, these are often used in combinations, however many of these agents cause severe side effects. Anthracyclines used in chemotherapy, such as Epirubicin, can cause cardiotoxicity; alkylating agents such as Cyclophosphamide, target hyperproliferating cells but are not specific to cancer cells and therefore can be toxic to somatic cells that are fast growing such as those in the gastrointestinal tract. There can be long term side effects when using these cytotoxic drugs and some treatments even increase the risk of contracting a secondary cancer such as Acute Myeloid Leukemia (AML) (Kaplan et al., 2007). A large percentage (around 80%) of patients receive some form of adjuvant therapy (Weigelt et al., 2005), such as chemotherapy, when only around half of these patients will relapse, and this can negatively affect their quality of life in the short term (hair loss,

sickness) and long term (cardiotoxicity, infertility) (Eifel et al., 2001) Thus there is a proportion of patients that would not suffer a relapse but would suffer unnecessarily as a consequence of adjuvant treatment. There is therefore an increasing requirement to identify predictive markers for breast cancer to avoid overtreatment and to better tailor treatment to individuals. There are few established prognostic biomarkers: the ERBB2 gene (HER2) biomarker has been extensively studied (Slamon et al., 1987; Reviewed by Ross et al.,2003) and more recently gene expression profiling has become popular as a way of predicting prognosis for individuals (Perou et al. 2000; Van't Veer et al., 2002).

5.1.4 HER2 status and breast cancer risk

The ERBB2 gene (encoding HER/neu) is overexpressed in 15-30% of breast cancer patients and has been correlated with poor prognosis; however the development of monoclonal antibodies to target the growth factor receptor has improved patient survival in HER2 positive (HER2+) patients. HER2 negative (HER2-) tumours fail to respond to monoclonal antibodies against HER2 receptors. Interestingly ERBB2 has been implicated as a potential predictive marker for therapeutic success (Houston et al., 1999). Breast tumours positive for ERBB2 expression have been shown to respond poorly to hormone therapies such as tamoxifen (Wright et al., 1992; Klijn et al., 1993; Tan and Yu, 2007), and this highlights the importance for distinguishing different breast cancer subtypes for more effective treatment.

5.1.5 Hormone receptor (HR) status and breast cancer risk: Oestrogen receptor (ER) and Progesterone receptor (PR) status as a biomarker

Oestrogen receptor (ER) positivity is found in a large proportion of breast cancer patients – as many as 60-70% (Rose et al., 1985; Walker et al., 2007; Wang et al., 2011). The development

of hormone therapies such as tamoxifen has enabled better treatment options for patients with ER positive (ER+) tumours. ER negative (ER-) tumours however tend to be more aggressive and are more likely to relapse and or metastasise (Schmidt et al., 1984; Crowe et al., 1991) and this is independent of demographic characteristics (Dunwald et al., 2006). Due to the insufficient knowledge attaining to the specific genetic differences between ER+ and ER-, there are less treatment options for ER- patients, this difficulty in treating these tumours may be one reason why ER- tumours have poorer prognosis than ER+ tumours.

As with ER positive breast cancers, progesterone receptor (PR) positive breast cancers respond well to hormone therapies (tamoxifen). Progesterone receptors exist in two isoforms: PR-A and PR-B. Studies have shown that patients with isoform PR-A are more likely to suffer a relapse when undergoing tamoxifen therapy (Hopp et al., 2004; Fuqua et al., 2005). Furthermore studies by Zhang et al. implicated low PR expression as having a role in an increase in growth factor signalling (i.e. increase in HER2 signals) and because of this an increase in the aggressive nature of the tumour (Zhang et al., 2005; Fuqua et al., 2005).

A number of deleterious prognostic implications can be drawn from these 3 biomarkers (HER2/ER/PR) when they are not expressed; tumours lacking these biomarkers tend to be more aggressive due to the lack of treatment options (Kaplan et al., 2006; Anders et al., 2009). Also, resistance against hormone therapies can occur in tumours overexpressing these markers, either *de novo* or developed, eg. trastuzumab or tamoxifen resisitance, rendering these breast cancers more aggressive and more difficult to treat (Nahta et al., 2006; Narayan et al., 2009). There is thus a requirement for further classification of breast cancer subtypes for a better understanding and thus more successful treatment.

5.1.6 The Nottingham prognostic Index (NPI) for prognosis

The Nottingham prognostic index (NPI) is used to determine prognosis following surgery for breast cancer (Haybittle et al., 1982; Elston et al., 1992). Its value is calculated using three pathological criteria: the size of the lesion; the number of involved lymph nodes; and the grade of the tumour (Todd et al., 1987; Elston et al., 1992). S is the size of the index lesion in centimetres, N is the number of lymph nodes involved: 0 =1, 1-3 = 2, >3 = 3 and G is the grade of tumour: Grade I =1, Grade II =2, Grade III =3 (Todd et al., 1987; Elston et al., 1992).

The index is calculated using the formula: NPI = [0.2 x S] + N + G. Scores calculated using this index are related to 5 year survival predictions (Table 5.1).

Table 5.1: NPI score related to 5 year survival predictions

Score	5-year survival
>/=2.0 to </=2.4	93%
>2.4 to </=3.4	85%
>3.4 to </=5.4	70%
>5.4	50%

Table 5.1: Shows the scores calculated for all NPI criteria using the formula NPI = [0.2 x S] + N + G and how they relate to five year predictions

5.1.7 Combined approaches for prognosis: NPI and Adjuvant!Online

Computer-based models are now being combined with clinical characteristics of breast tumours, giving algorithms to predict risk of relapse (Geyer et al., 2009). One example of this is Adjuvant!Online whereby patient information such as age, clinical characteristics of tumour as determined by NPI, ER/PR status and lymph nodes involvement. The relative risk of survival and relapse at 10 years of (without additional therapy) is calculated as a guide for patient treatment options (Gribbon and Dewis, 2009) i.e. a guide as to whether or not oncologists should administer chemotherapy to a patient.

5.1.8 Gene expression based techiques- Mammaprint, OncotypeDX

High throughput molecular classification techniques are being used with increasing popularity. Collections of gene from different prognostic groups, such as ER+, are used as predictive signatures to further classify independent groups. The expression level of key genes by microarray analysis is a novel multivariate analysis to predict distant metastases. The first recognised prognostic signature used that compares gene expression profiles of breast tumours in both lymph node positive and negative tumours was developed by Agendia and is called MammaPrint (Van't Veer et al 2002 and 2006). This technique uses a 70 gene signature to distinguish between patients that are likely to develop distant metastases after 5 years and those that are not. Variations on this microarray based prognostic technique have been developed such as a 21 gene assay by OncoTypeDX, using 16 breast cancer-related genes and 5 normal 'reference' genes to give a recurrence score (RS) (a high RS denotes a poor prognosis) (Paik et al., 2004). These techniques have merit and MammaPrint is FDA approved and undergoing MINIDACT (Microarray In Node-negative Disease may Avoid Chemotherapy) patient trials. OncoTypeDX has been approved for use by the American Society of Clinical Oncology (ASCO). There are drawbacks with these technologies.

MammaPrint has limited efficacy for ER- tumours (Geyer et al., 2009) and using OncoTypeDX gene expression profiles are assessed for ER positive, lymph node negative breast tumours only (Paik et al., 2004).

5.1.9 NGS sequencing in breast cancer- defining multiple subtypes with differential risks

To further define molecular subtypes Curtis et al. utilised next generation sequencing (NGS) in a large cohort study. Next generation sequencing allows sequencing-by-synthesis reactions to be run in parallel in order to provide high throughput sequencing. This allowed copy number and gene expression to be analysed in 2,000 breast tumours (Curtis et al., 2012). Breast tumours are driven by genomic instabilities and changes in copy number (CNA-copy number alterations). However there exist large degrees of somatic (non-oncogenic) copy number variation (CNV) and single nucleotide polymorphisms (SNPs). The presences of CNAs, CNVs and SNPs in breast cancer genome were mapped by using NGS technology and this was integrated alongside gene expression to elucidate clearer breast cancer subgroups based on clinical outcome. This technique gives a method of understanding how CNAs affect gene expression profiles which could help to understand how different breast cancer types respond differently to therapies. Furthermore certain high risk subgroups were defined using this technique and this could be a starting point for future studies (Curtis et al., 2012).

5.1.10 Limitations of these risk categories

Therapies tend to be tested in specific subgroups, for example lymph node negative ER+ patients with OncoTypeDX, instead of assessing risks across all individual patients. Defining risk at the individual level is more difficult therefore there is a requirement for more

comprehensive prognostic information.

5.1.11: Kaplan-Meier analysis of common biomarkers in Breast Cancer

Multivariate analysis of common breast cancer biomarkers was performed to elucidate the prognostic significance over a long follow up period (Hilsebeck et al., 1998). Cox proportaional hazards models coupled with diagnostic plots were evaluated for several factors: including tumour size, axillary lymph nodes and estrogen receptor (ER) status. This was completed for a large sample of breast cancer cases (N=2,873) with up to 17 years of follow-up for disease-free survival (DFS). Presence of positive nodes (axillary nodes) was found to be the most significant predictor of survival outcome with a hazard ratio of 1.86 ($P<0.0001$).

5.1.12 Telomere length as a prognostic marker

Few studies have shown a direct statistical correlation between telomere length and prognosis. Correlations have been shown between telomere length and severity of disease, for example studies using paraffin embedded invasive human breast cancer samples have showed that tumours with the shortest telomere length had the greatest amount of metastasis (significance $p < 0.05$) (Heaphy et al., 2007). However no direct link has been used to determine whether telomere length can be used to predict survival in a statistically significant manner.

5.1.13 Project Aims

Telomere lengths analysed by STELA, a high resolution single molecule strategy will be compared with clinical data received for all samples. Clinical data collected over a 5 year follow up period includes information about: receptor status, lymph node involvement, grade, morbidity, whether or not the patient has relapsed and HER2 status. Kaplan Meier analysis will allow the calculation of hazard ratios pertaining to survival statistics. By linking the clinical data with the calculated telomere lengths it is hoped conclusions can be drawn about telomere length as a potential prognostic indicator.

5.2 Results:

To evaluate high resolution telomere length data as a prognostic marker in breast cancer clinical data was obtained from 120 patients samples from the Wales Cancer Bank over a 5 year follow up period. This data included information about survival, grade, lymph node status, relapse and age.

The mean age of the cohort was 62 years old.

For overall survival only deaths that were confirmed as related to breast cancer were included in the analysis (n=120) (Figure 5.1).

Figure 5.1: Kaplan-Meier Curve for whole cohort

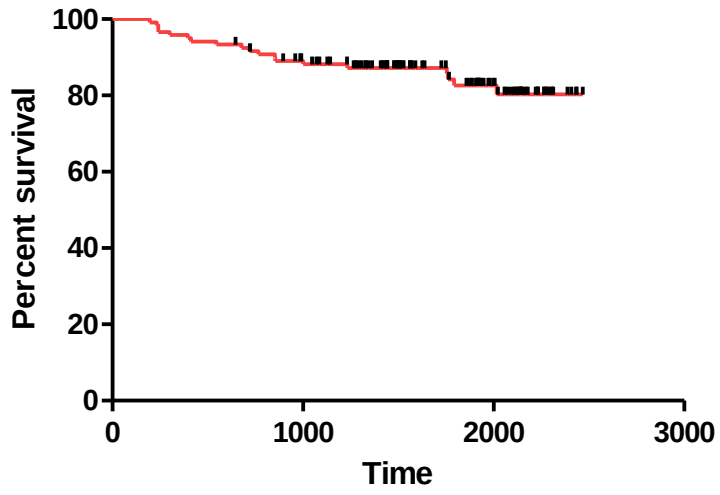


Figure 5.1: Kaplan-Meier curve including entire patient cohort with a total of 19 deaths.

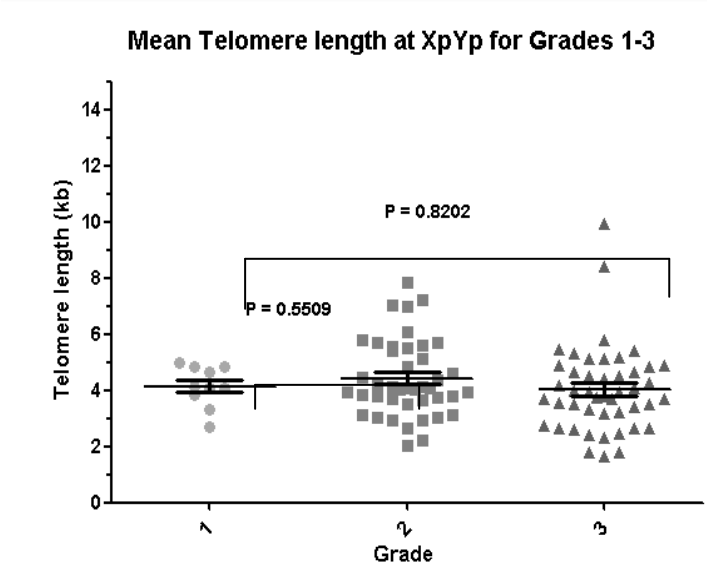
5.3 Relation between XpYp telomere length profiles and histological grade

Mean telomere length and standard deviation at XpYp were calculated for each patient sample and grouped by grade. Grade 1 cells have the most normal cellular histology, and grade 3 cells have the most clonal and fast growing cellular phenotype. Plotting these values

on an XY scatter plot (figure 5.2a), revealed no significant difference between the telomere lengths in each subgroup (p=0.4; Figure 5.2a). Given the relationship between telomere length heterogeneity and clonal composition, standard deviations were also compared to grade. No significant correlation was found (p=0.6, figure 5.2b).

Included in the plot containing SDs for each histological grade were the SDs that were obtained from single cell clones (Baird et al. 2003). This was plotted alongside the IDC data to illustrate that some of the tumours do have relatively homogenous distributions that are consistent with clonal growth. Quantification of the proportion of tumours with telomeres in this range revealed that 20% of Grade I tumours were in the clonal range; 16% of Grade II and 30% of Grade III (giving an average of 22% across all grades). Telomeres in this range are shown below the red dotted line shown in figure 5.2b.

Figure 5.2a: Mean telomere length scatter plot for Grades 1-3 at XpYp



One-way analysis of variance	
P value	0.4402
P value summary	ns
Are means signif. different? (P < 0.05)	No
Number of groups	3
F	0.8277
R squared	0.01713
Bartlett's test for equal variances	
Bartlett's statistic (corrected)	6.041
P value	0.0488
P value summary	*
Do the variances differ signif. (P < 0.05)	Yes

Figure 5.2a: Figure shows a scatter plot of mean telomere lengths for each histological grade. Grade 1 is classed as the least severe subtype and grade 3 the worst in terms of prognosis. The P-values shown on the graph were from Student's t-test to compare telomere means.

Figure 5.2b: Standard deviation scatter plot for Grades 1-3 at XpYp

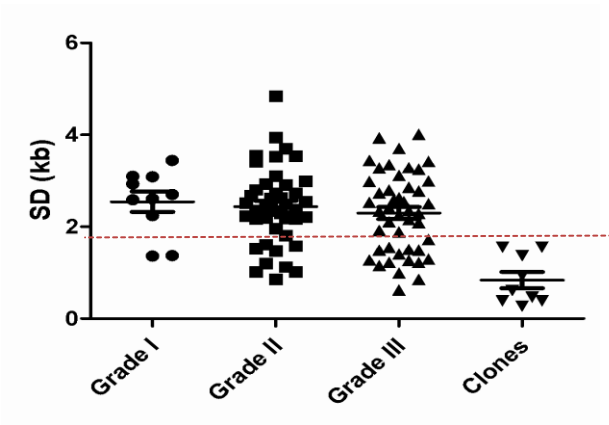


Figure 5.2b: Figure shows a scatter plot of standard deviations (S.D.) for each histological

grade. Bartlett's test for equal variances was not significant p=0.62. Alongside the S.D. values observed in the IDC samples is a collection of S.Ds obtained from single cell clones (from Baird et al., 2003) to demonstrate the percentage of IDC samples in this range (the upper standard deviation limit calculated for telomere lengths analysed in single cell clones is illustrated by a red dotted line to show the patients that fall in this range from the IDC cohort)

5.4 Assessing the performance of the commonly used prognostic markers in IDC

Kaplan-Meier analysis is a way of relating survival data, using survival curves, with prognostic factors such as the presence of hormone receptors. Kaplan-Meier statistics generate a hazard ratio which is a predictive value of the survival outcome i.e. a ratio of four would mean a patient with the presence of such factor would be four times more likely to die of the disease in unit time; a hazard ratio of one means that there is no difference in survival.

5.4.1: ER, PR and HER2 status

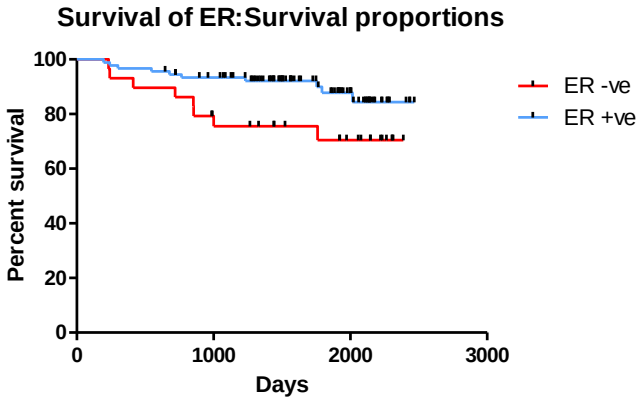
Firstly patients were grouped into two groups: ER positive and ER negative. Kaplan-Meier analysis showed that patients with ER positive subtypes had poorer prognosis than those without, with a hazard ratio of 3.096 (p=0.02). The survival curves were significantly different (figure 5.13).

Partitioning patients based on progesterone receptor (PR) status, i.e. positive and negative. Kaplan-Meier survival curves generated for this clinical information were significantly

different and gave a hazard ratio of 2.798 (p=0.0364) (figure 5.14).

HER2 positive patients were initially compared directly with HER2 negative patients for Kaplan-Meier survival analysis. The hazard ratio generated for this analysis was 1.549 and the survival curves were not significantly different (P=0.4513 Mantel-Cox test; figure 5.15).

Figure 5.13: Kaplan-Meier of Survival vs. ER positivity



Hazard Ratio	
Ratio	3.096
95% CI of ratio	1.047 to 9.157

Log-rank (Mantel-Cox) Test	
Chi square	4.173
df	1
P value	0.0411
P value summary	*
Are the survival curves sig different?	Yes

Figure 5.13: Kaplan-Meier curves showing the survival pertaining to Oestrogen receptor (ER) status showing ER negative patients having a predicted poorer prognostic outcome (HR 3.096; p=0.0411)

Figure 5.14: Kaplan-Meier of Survival vs. PGR positivity

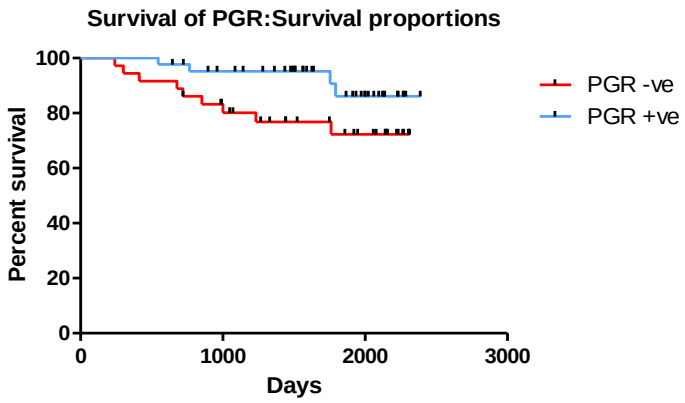


Figure 5.14: Kaplan-Meier curves showing the survival pertaining to Progesterone receptor (PGR) status showing PGR negative patients having a predicted poorer prognostic outcome (HR=2.798; p=0.0364)

Figure 5.15: Kaplan-Meier of Survival vs. HER positivity

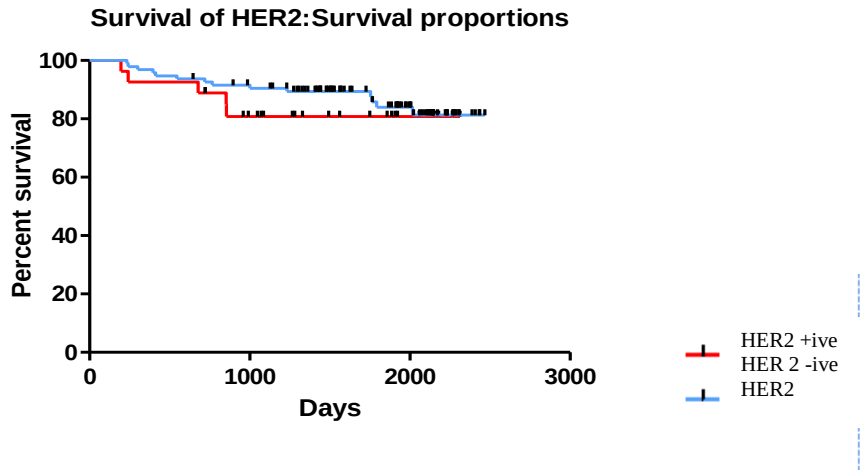


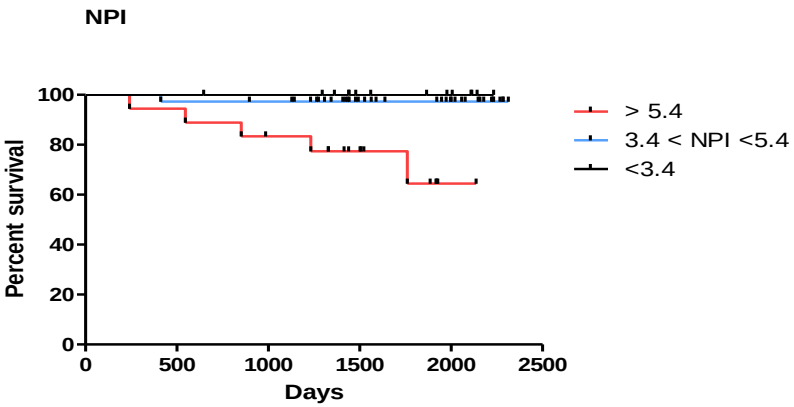
Figure 5.15: Kaplan-Meier curves showing the survival pertaining to HER2 status. There is no significant difference in the survival curves when using HER2 status as a prognostic factor. (HR=1.549; P=0.4513 Mantel-Cox test)

5.4.2: Kaplan-Meier statistics for Nottingham prognostic index and telomere length

The Nottingham prognostic index (NPI) is used to determine prognosis following surgery for breast cancer. Kaplan-Meier analysis was applied to clinical data collected over a five year period for NPI for each patient. The Nottingham Grading system for breast cancer combines three main histological features: nuclear grade, tubule formation and mitotic rate. Each factor is given a score of 1 to 3, 3 being most severe, for example cells that are most abnormal in appearance and mitotic rate would be scored 3. The score of all three components are added together to give the "grade". A score of 3-5 is considered low grade (I); a score of 6-7 is graded intermediate (II) and 8-9 would be classified as high grade (III).

Patients were partitioned by NPI values into three groups: <3.4, 3.4-5.4 and >5.5. The resulting survival curves show a significant difference (P=0.0033) between each group, with a lower NPI score relating to better survival outcome over time (figure 5.16).

Figure 5.16: Kaplan-Meier of Survival vs. NPI



Log-rank (Mantel-Cox) Test	
Chi square	11.65
df	2
P value	0.0030
P value summary	**
Are the survival curves sig different?	Yes

Figure 5.16: Partitioning patients by NPI values show a significant difference (P=0.0033) between each group, with a lower NPI score relating to better survival outcome over time

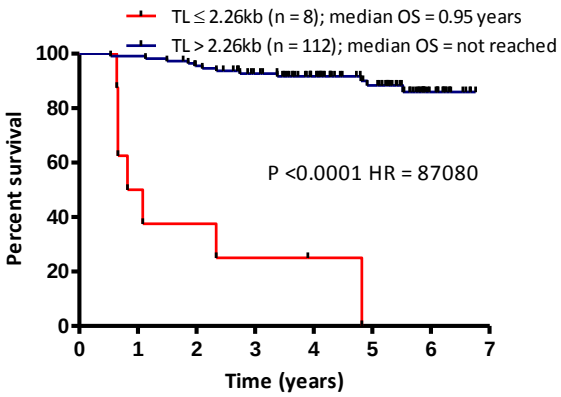
5.5: Telomere length (XpYp) vs survival using Kaplan-Meier curves

Telomere length data collected for XpYp was correlated with survival statistically using Kaplan-Meier survival curves. Original means calculated from STELA blots and Gaussian adjusted means were used for Kaplan-Meier analysis due to the presence of bimodal telomere length distributions (chapter 4-telomere length analysis). Comparisons of the two mean groups for survival analyses were almost identical in outcome (P<0.001). It could be concluded that removing a few outliers from mean distributions using Gaussian statistics was not affecting the overall means from the dataset (data not shown).

As several tumours exhibited bimodal telomere-length distributions as discussed in chapter 4 STELA profiles were thus reanalysed so that only telomere lengths in the lower bimodal group where used for the survival analysis. This would thus account for any infiltrating telomere lengths from other cell types and just measure the mean for the lowest clusters of

telomeres that are hypothesised to be more closely reminiscent of the tumour cell types. Results from Kaplan-Meier curves showed that telomere lengths below 2.26 kb had the most significant ($p < 0.0001$) negative impact on survival (figure 5.17; patients had a hazard ratio of 87080 and thus are 87080 times more likely to die from invasive ductal carcinoma if they have a mean telomere length that is below 2.26kb. The hazard ratios calculated for this data set are the highest hazard ratios that have been documented for breast cancer survival. The data clearly identifies a subset (7%) of patients with IDC that exhibit a considerably reduced survival (median survival 0.95 years).

Figure 5.17: Kaplan-Meier analysis curves for data recursively partitioned by 2.26kb



Gehan-Breslow-Wilcoxon Test	
Chi square	67.14
df	1
P value	< 0.0001
P value summary	***
Are the survival curves sig different?	Yes
Median survival	
≤2.26kb	0.9534
>2.26kb	Undefined
Hazard Ratio	
Ratio	87080
95% CI of ratio	6398 to 1185000

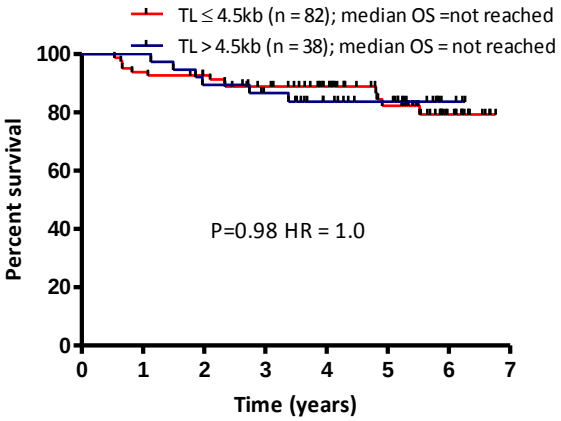
Figure 5.17: Patients partitioned by telomere length 2.26 kb gave significant prognostic outcome for survival (HR= 87080; p<0.0001)

5.6 Recursive partitioning

To establish if the 2.26 kb represented the optimal threshold for prognostication in breast cancer recursive partitioning was performed. The mean telomere length of the whole cohort was 4 kb, so this was used as a starting point for partitioning the data. Means above and below this value were also used, from 4.5 kb to 1.79 kb (lower quartile (Figure 5.18a and b). Partitioning data by the higher means showed low hazard ratios of below 1 and no significant difference in curves (P=0.6 for TL 4.0kb; p=0.8 for TL 4.5kb). As the length of the telomere mean used to partition the data decreased the hazard ratio began to increase gradually: at TL 3.0 kb the hazard ratio was 3.4 (p=0.03) rising to over 87,000 at 2.26 kb (figure 5.19).

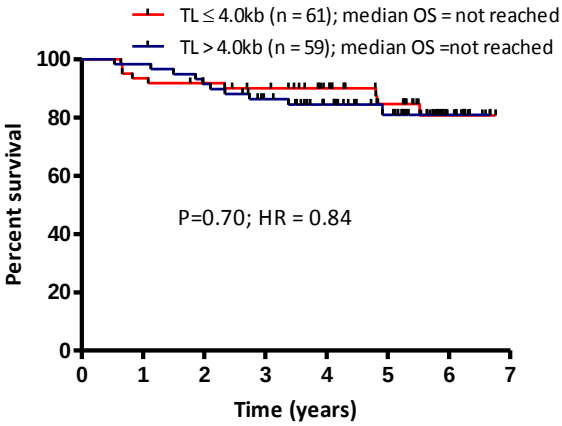
It was clear that the threshold telomere length below which telomeres have been shown to readily undergo fusionsis most prognostic in IDC.

Figure 5.18a : Kaplan-Meier curves for TL 4.5 and 4.0



Gehan-Breslow-Wilcoxon Test	
Chi square	0.05178
df	1
P value	0.8200
P value summary	ns
Are the survival curves sig different?	No

Hazard Ratio	
Ratio	0.9874
95% CI of ratio	0.3739 to 2.607

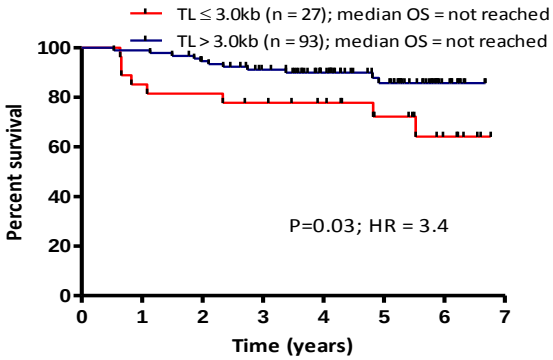


Gehan-Breslow-Wilcoxon Test	
Chi square	0.2518
df	1
P value	0.6158
P value summary	ns
Are the survival curves sig different?	No

Hazard Ratio	
Ratio	0.8370
95% CI of ratio	0.3395 to 2.063

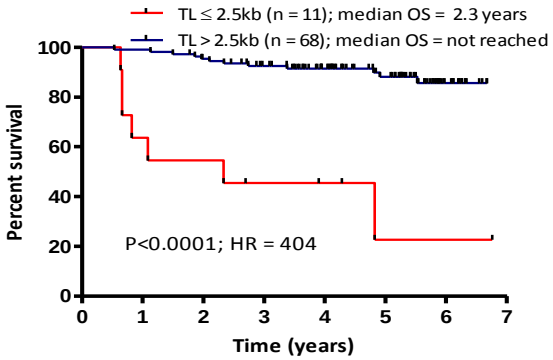
Figure 5.18a: All Kaplan-Meier plots from rescursive partitioning by telomere lengths of 4.5kb and 4.0kb gave low hazard ratios of 0.9874 (p=0.8) and 0.8370 (p=0.6) respectively. Curves showed no significant difference when partitioning data by these telomere length means.

Figure 5.18b: Kaplan-Meier curves for TL 3.0 and 2.5



Gehan-Breslow-Wilcoxon Test	
Chi square	4.693
df	1
P value	0.0303
P value summary	*
Are the survival curves sig different?	Yes

Hazard Ratio	
Ratio	3.428
95% CI of ratio	1.144 to 10.27



Gehan-Breslow-Wilcoxon Test	
Chi square	35.34
df	1
P value	< 0.0001
P value summary	***
Are the survival curves sig different?	Yes

Hazard Ratio	
Ratio	404.4
95% CI of ratio	56.40 to 2900

Figure 5.18b: Rescursive partitioning by shorter telomere lengths of 3.0kb and 2.5kb gave higher hazard ratios (3.4 and 404.4 respectively) with significantly different curves (TL of 3kb p=0.03 and TL of 2.5 p<0.0001) showing the shorter the telomere the more significant the curve separation.

Figure 5.19 Summary HR plot that confirms the 2.26 kb threshold:

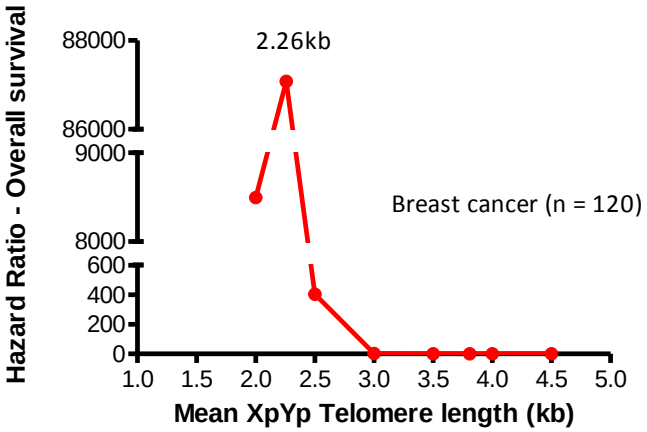


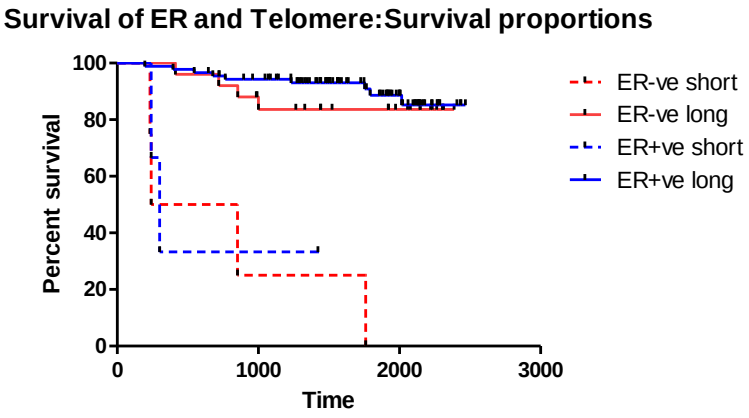
Figure 5.19: Plot of all hazard ratios calculated following the stratification of patients based on the mean XpYp telomere length as indicated on the X-axis.

5.7 Telomere length splits prognostic subsets based on NPI, ER, HER2 and PGR status

5.7.1 ER:

As mentioned previously correlations between ER status and survival using Kaplan-Meier plots resulted in the calculation of a hazard ratio of 3.096 (p=0.02) (figure 5.13). The combined analysis of telomere length with estrogen receptor status showed that telomere length was able to split the good and bad prognostic ER subgroups. Kaplan-Meier curves showed that ER positive and ER negative patients had the worst survival outcome if their telomeres were below 2.26kb, and patients with longer telomeres had fewer deaths over time (both ER positive and ER negative) (figure 5.20). This data indicate that telomere length is independent of ER status for prognosis in IDC.

Figure 5.20: Kaplan-Meier of ER positivity and Telomere length vs. Survival



Log-rank (Mantel-Cox) Test	
Chi square	58.93
df	3
P value	< 0.0001
P value summary	***
Are the survival curves sig different?	Yes

Figure 5.20: Multivariate analysis using ER status and the prognostic telomere length threshold of 2.26kb highlights the prognostic power of telomere length independent of steroid hormone receptor status. Both ER positive (blue) and ER negative (red) have a worse prognosis with telomere lengths below 2.26kb (shown on graph as dashed lines).

5.7.2 PR status

The patient data assessed previously using Kaplan-Meier curves was again grouped by telomere length above and below the threshold of 2.26kb. None of the patients with short telomeres were PR positive. However, regardless of PR status, the patients with short telomeres and negative PR status had a worse predicted survival outcome than PR positive patients with long telomere profiles (Figure 5.21). Comparison of PR positive patients with long telomeres to PR negative patients with short telomeres gave a hazard ratio of 21900.

This is a much higher hazard ratio than was calculated for PR status alone.

Figure 5.21: Kaplan-Meier of Survival vs. PGR positivity and Telomere length

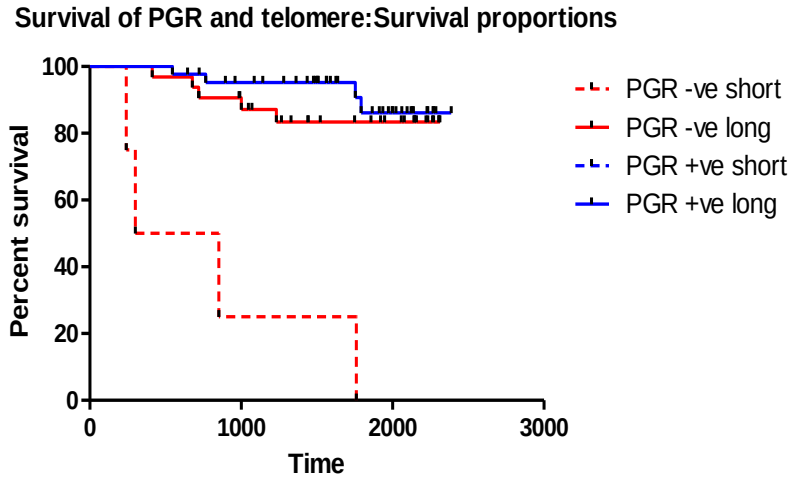


Figure 5.21: Multivariate analysis using PGR status and the prognostic telomere length threshold of 2.26kb highlights the prognostic power of telomere length independent of steroid hormone receptor status again. Both PR positive (blue) and PR negative (red) have a less severe prognosis with telomere lengths above 2.26kb (therefore 'long')(shown on graph as dashed lines), however no patients fell into the category of PGR positive and short.

5.7.3 HER2 status

On factoring in the telomere length profiles for each patient and partitioning HER2 status by a telomere length profile of above and below 2.26kb, the survival curves changed dramatically. As seen with the other subtypes, HER2 status was not the defining factor, but all patients with short telomere length profiles had a worse survival prediction over time. Kaplan-Meier curves generated for this analysis are shown in figure 5.22. However these observations should be taken with the caveat that the two inferior curves contained just 2-3 patients.

Figure 5.22: Survival of patients positive for HER2 vs patients with short Telomere length

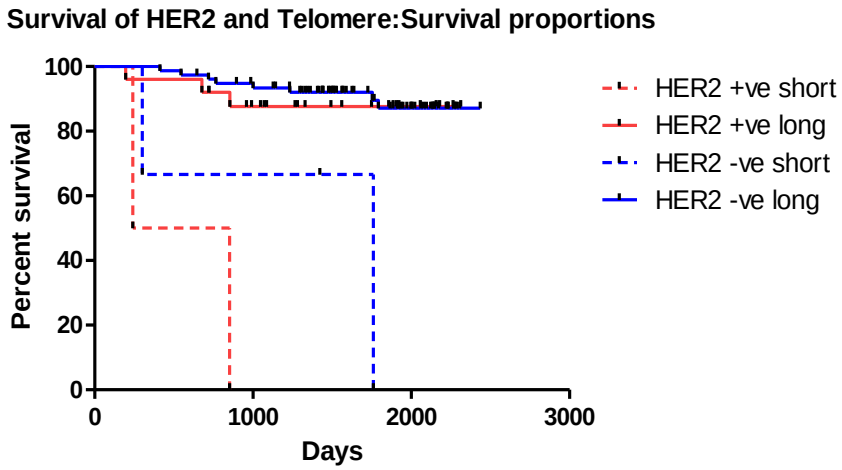
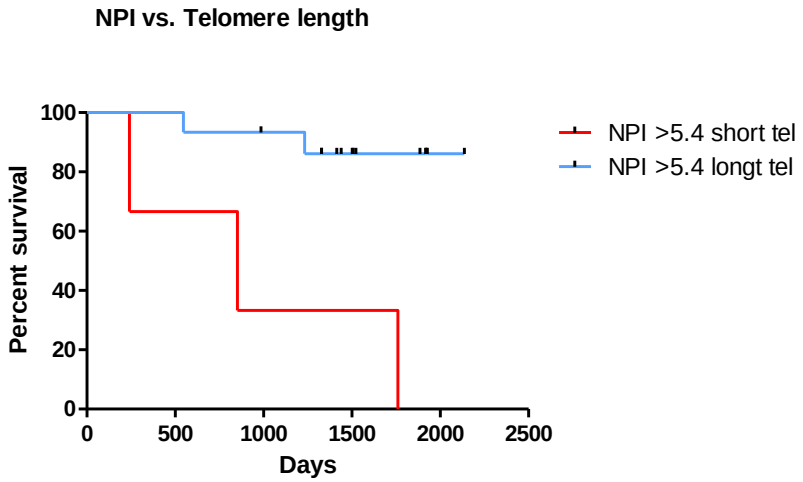


Figure 5.22: HER2 status was partitioned by telomere length. Both HER2 positive (blue) and HER negative (red) have a worse prognosis with telomere lengths below 2.26kb (shown on graph as dashed lines), however there is only a small number of patients in this grouping.

5.7.4 NPI

Combined analysis of telomere length with NPI was also carried out. Telomeres in each group (short/long) were categorised by a score of above 5.4 as this was shown to have the biggest impact on survival. Findings reveal that IDCs with short telomeres and an NPI score of >5.4 had a worse prognosis than those with long telomeres (figure 5.23). The hazard ratio calculated from this was 68.14, which is extremely high compared to hazard ratios calculated previously for breast cancer. This means that patients with a score of above 5.4 and short telomeres were 68 times more likely to die of the disease than those with long telomeres and the same NPI scores. This again shows the importance of telomere length as a prognostic predictive factor for breast cancer survival outcome.

Figure 5.23: Kaplan-Meier of Survival vs. NPI partitioned by Telomere Length



Hazard Ratio	
Ratio	68.14
95% CI of ratio	4.933 to 941.1

Figure 5.23: Partitioning patients by telomere length proved more significant than the worst NPI score (HR=68.14, p=0.0016)

5.8 Grade

XpYp telomere lengths grouped as long and short (above and below 2.26kb) were compared with survival in groups II and III. Comparing patients with short telomeres in group III gave a hazard ratio of 226 ($p<0.0001$; figure 5.24), i.e. patients with a short telomere length profiles are more likely to die from the invasive ductal carcinoma than patients with long telomeres in the same grade group. A similar and striking observation was also found with data for patients in the grade II group. The hazard ratio calculated comparing telomere lengths in this grouping was over $7x10^{13}$, however the inferior curve contained just 2 individuals (figure 5.25).

Figure 5.24 Patients partitioned by Grade III for TL 2.26kb

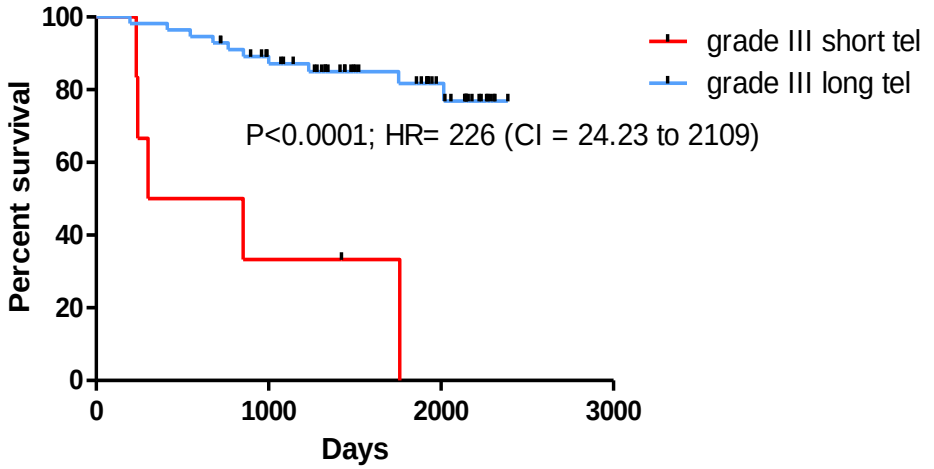


Figure 5.24: Patients displayed on survival curve are all grade III for NPI score and partitioned by telomere lengths of 2.26kb i.e. above 2.26kb classified as 'long'. Curve comparison is statistically significant (P<0.0001) and telomere lengths below 2.26kb are prognostic with a hazard ratio of 226.

Figure 5.25 Patients partitioned by Grade II for TL 2.26kb

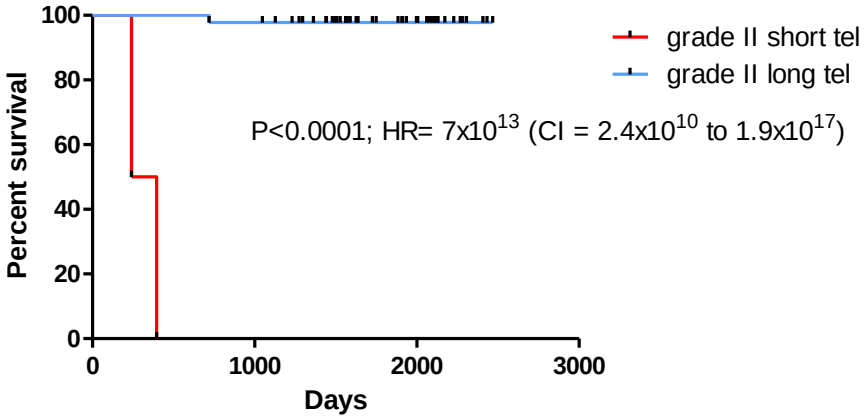


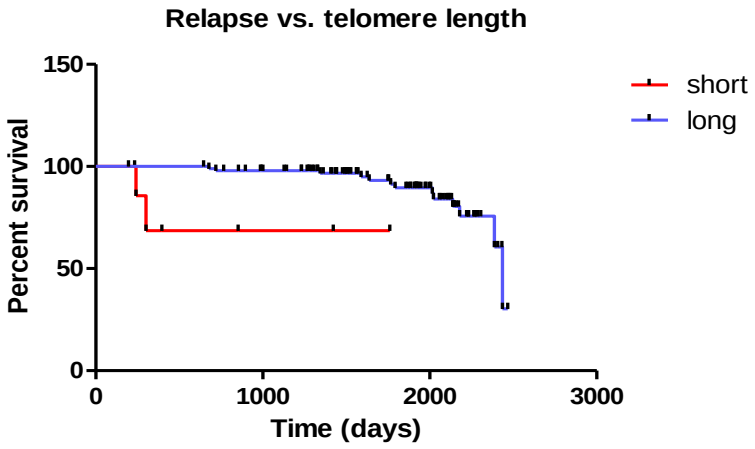
Figure 5.25: Patients displayed on survival curve are all grade II for NPI score and partitioned by telomere lengths of 2.26kb i.e. above 2.26kb classified as 'long'. Curve comparison is statistically significant (P<0.0001) however only 2 individuals are represented on the inferior curve and thus the high hazard ratio may not be an adequate representation.

5.8: Kaplan-Meier survival analysis for Relapse verses Telomere Length

The risk of relapse is much higher triple negative breast carcinomas than hormone related subtypes for the first 5 years. After this time point, research shows the risk of relapse decreases below that of hormone-positive breast carcinomas.

Clinical information about whether or not there has been a relapse was compared with the telomere length profiles for each patient and was analysed by Kaplan-Meier statistics. Relapse data pertains to any patient that has had a recurrence i.e. cancer cell growth that arises after initial treatment/detection. Recurrences can occur months or even years after the initial treatment, and either in the same place as the original tumour site ('local recurrence') or distant metastases including bones, skin, liver, lungs or brain. Results from input of this information gave a hazard ratio of 2252 (p value=0.0002) (figure 5.26), showing that patients with short telomeres were more likely to relapse than those with longer telomeres.

Figure 5.26 Survival for Relapse vs. Telomere length



Hazard Ratio	
Ratio	2252
95% CI of ratio	37.21 to 136300

Figure 5.26: Patients in the 'short' telomere category have telomeres of 2.26 and below and show a poorer prognosis in terms of relapse HR= 2252; p value=0.0002.

5.9.1: Discussion: Telomere length as a prognostic factor

The most striking observation from the clinical of analysis of this breast cancer cohort analysed for telomere length at XpYp is that stratification of patients based on the threshold telomere length at which fusion occurs (2.26kb) reveals a remarkable level of prognostic power for overall survival. Breast cancer is a heterogeneous disease and there are some breast tumour types and are more likely to relapse/ metastasise (Schmidt et al., 1984; Crowe et al., 1991). To prevent this, a large number of patients receive some form of adjuvant therapy, many of which can be detrimental to quality of life and even cause the patient to develop leukaemia (Kaplan et al., 2007). Biomarker discovery is important to provide more prognosis of breast cancer patients as the standard methods (e.g. NPI, hormone status) do not give a clear prediction for relapse and patients are often over treated as a result (Weigelt et al., 2005). The development of microarray techniques, such as MammaPrint (Van't Veer et al., 2005), has allowed gene expression profiles to be used as a way to assess breast cancer subtypes. These techniques are becoming more widely used and are being approved for patient trials, however these gene expression profiles tend to be tested in specific subgroups, such as ER positive, and could not be used for all individuals and provide little information on overall survival. Newer NGS technology allows the sequencing of entire breast cancer genomes to be paired with microarray data to give a better definition of breast cancer subtypes (Curtis et al., 2012).

5.9.2 Telomere length and prognosis in the literature

A few studies have correlated telomere length with breast cancer severity, for example in a study by Heaphy *et al.*, mean telomere lengths for all chromosomes were assessed in four different mammary tumour subtypes: luminal A, luminal B, HER2 and triple negative (Heaphy et al., 2010). Luminal A tumours tend to be less aggressive than luminal B subtypes

and respond better to oestrogen receptor targeting. Triple negative tumours and tumours overexpressing HER2 tend to convey poor prognosis. Findings from applying telomere *FISH* to detect telomere lengths in each subtype of mammary cancer revealed a mean shorter telomere length in the more aggressive subtypes: luminal B, HER2 and triple negative. Longer mean telomere lengths were recorded for luminal A carcinomas. This research shows a correlation between telomere shortening and established prognostic subtypes of breast cancer. Telomere lengths were qualitatively scored by direct visual assessment of stained slides (*FISH)* and this method is very subjective and relies on telomeres that are long enough to be detected by fluorescent probes. This research does point towards telomere length as a useful tool for determining prognostic outcome however using STELA for higher resolution of telomere length quantification in different tumour subtypes has already shown an extremely significant outcome when calculating survival proportions.

5.9.3 Telomere length as a biomarker

Numerous studies have looked at common biomarkers, such as Her2/ER, that confer risk or hazards for breast cancer related unfavourable events i.e., death due to breast cancer, breast cancer recurrence, or development of a new primary breast tumour. To date, not many of studies have looked at telomere length in relation to breast cancer risk for survival and local recurrence. A study by Zhou *et al.* aimed to investigate whether telomere length in normal breast epithelial cells surrounding the tumour is predictive of breast cancer local recurrence (LR) (Zhou et al., 2011). Previous studies have shown that very short telomere length is a common genetic abnormality in pre-malignant breast lesions and early breast cancer cells. Results from using *FISH* to detect telomere lengths in samples showed that smaller telomere length variation (TLV) in normal epithelial cells adjacent to the tumour had a significantly increased risk of LR. More heterogeneous distributions were associated with significantly better 10-year recurrence free survival. TLV in normal epithelial cells surrounding the tumour

was more significantly associated with LR risk compared with tumour cells. This is evidence towards the hypothesis that LR after surgery to remove tumour tissue can be in part due to failure to remove sufficient pre-cancerous cells in the vicinity of the tumour bed margins. Telomere length variation studies could reflect the recent cell proliferative history. This study had flaws however, the sample size was relatively small and the distance between normal epithelium and tumour cells was unknown.

Experiments looking at hazard risk for survival have shown that telomere content (TC) determined by slot blot analysis, predicts breast cancer–free survival, independent of other clinical and prognostic factors (Griffith, *et al.,* 2006; Heaphy, *et al.,* 2007). Experiments by Heaphy *et al.* showed that tumours with a low TC content compared with a placental DNA standard conferred a hazard ratio for overall breast cancer survival as 2.25 (95% CI, 1.09-4.64; P = 0.029). Another study by Griffiths *et al.,* showed that low TC conferred an unadjusted relative hazard of 4.39 (95% CI=1.47–13.08; p=0.008) relative to high TC. These hazard ratios are low compared with the hazard ratios calculated from telomere length data generated by STELA. STELA data has not only provided the highest recorded hazard ratios documented for breast cancer, but appears to be a more powerful biomarker than current biomarkers used such as HER2 and ER status. This is because telomere length analysis is applicable to all patients and is thus a universal prognostic marker that is highly prognostic independent on the presence of cell receptor status. This analysis was carried out in a small cohort, however even though only 7% of patients were in the short telomere category, statistically comparing telomere length as a survival factor against common biomarkers revealed extremely high hazard ratios and thus this need to be tested in larger cohorts for validation. In a review of 107 studies, assessing around 40,000 patients, HER2 status was documented as having a relative risk of 2.74 (CI: 1.39-6.93) for poor prognosis (Ross et al., 2009) which is similar to the hazard ratio attained from the clinical data received for this

dataset for IDC, but on a much larger scale. Telomere length however greatly exceeds this value. It is thus fair to conclude that telomere length has a much higher prognostic power than ER, HER2 and NPI status. Telomere length has already shown to be highly prognostic in a completely different type of cancer (CLL; Lin et al 2010) with the exact same prognostic threshold length. It could thus be speculated that this threshold length is a biological constant below which telomere dysfunction occurs.

5.9.4 Telomere length as a biomarker in human disease

Abnormal telomere lengths have been documented as being a contributing factor in many pathological conditions, not only cancer but age-related diseases and premature ageing syndromes [Reviewed by Blasco, 2005]. Telomere length is regulated epigenetically by changes to chromatin structures (Ahmad & Henikoff, 2002; Garcia Cao et al., 2004.) Thus diseases that have epigenetic defects may have abnormal telomere length regulation, such as patients with Rett's syndrome that have mutations in MECP2 (Ahmad & Henikoff, 2002; Blasco, 2005).

TRF2 protects the G-rich overhang at the telomere terminus and interacts with repair proteins such as ATM (Smogorzewska and de Lange, 2004; Karlseder. et al., 2004). Links have been shown between defects in DNA damage repair and short telomeres as having a causal role in some diseases such a Nijmegen break syndrome (Ranganathan et al., 2001). TRF2 loss has been shown to cause an increase in telomere fusions (van Steensel, 1998), and components that associate with TRF2 such as NSB1 are implicated in the genomic instabilities in diseases (Ranganathan et al., 2001), thus telomere fusion and defective DNA damage repair may be driving genomic instabilities that give rise to a whole host of diseases. Other diseases linked to defective DNA damage repair and genomic instability such as Fanconi Aneamia have been correlated with short telomere length (Callén et al., 2002). A large percentage of patients with Fanconi Aneamia go on to develop haematological cancers [Reviewed by Alter, 2003], thus

the use of STELA as a biomarker could reflect prognostic severity of condition and may be a way of predicting clinical outcome in patients with this condition. Short telomeres have been correlative with poor prognosis in patients with heart disease and infections and thus it is not onlyage-related diseases and cancer that may have some type of telomere dysfunction (Cawthon et al., 2003).

Thus using STELA as a high resolution technology to assess telomere length analysis may be applicable to a whole host of human disorders, and not just cancer prognosis.

5.9.5 Telomere length is a driving force behind progression to malignancy

Studies in vivo have shown the presence of extremely short telomeres across all Binet's stages of CLL, getting progressively shorter with poorer prognostic groups. Telomere erosion to lengths at which cells are in crisis, in ranges that have been shown to display high frequencies of telomere fusions in vitro have been observed in CLL and correlated with an increase in genomic instability. Furthermore some earlier stage CLL patients also exhibited telomeres in the dysfunctional length range (Lin et al., 2010). Telomere erosion is thus considered to have a causative role in the genomic instabilities that drive progression to malignancy. In this breast cancer cohort (IDC) short telomeres within the dysfunctional range have been observed across all NPI grades and seem to be independent of receptor status (HER2/ER), although with a small cohort it is difficult to draw significant conclusions. It would be interesting to expand this cohort with a variation of different subtypes, not just differing in receptor status but in invasive vs. *in situ* to assess telomere lengths across subtypes. I would expect that telomeres of critical lengths would be evident in all subtypes, but perhaps more so in the later stages. It is possible that in breast cancer telomere erosion is an early event, acting as a driving force behind progression to malignancy. Telomere erosion has been shown to be facilitated by the inability of DSB checkpoints to recognise dysfunction allowing cells to divide past crisis both in vitro with the abrogation of p53 (Lin et al., 2010)

and in CLL samples (Britt-Compton et al., 2012). LOH studies in breast cancer have shown the loss of p53 and ATM, in the earliest stages of breast cancer (Shen et al., 2000) which supports the hypothesis that short telomeres in the absence of functional DNA damage responses are driving chromosomal instabilities and the progression to malignancy.

Telomerase is found to be overexpressed in the majority of human cancers (Shay and Bachetti, 1997). Telomerase activation must occur in later stages of cancer as a way of preventing mitotic catastrophe. TERT expression in AML illustrates this process whereby clonal expansion of leukemic cells leads to telomere shortening and then is reactivated to avoid cellular senescence (Hartmann et al., 2005). Some later NPI grades have longer telomeres, and this could be evidence that there is a two-step process at work – first, in the absence of telomerase telomeres get shorten and chromosomes fuse together. Then telomerase is reactivated rendering cancer cells immortal. One way to verify this would be to apply the TRAP (Telomeric repeat amplification protocol) assay to the breast cohort to assess the role of telomerase in the mutagenic process. Studies in telomerase negative fibroblast cell lines have shown that the shortest telomeres are preferentially elongated in human cells (Britt-Compton et al., 2009), therefore it would be interesting to assess the expression of telomerase in the breast cancer cohort. In 15% of human cancers an alternative telomere lengthening mechanism (ALT) has been documented (Bryant et al., 1997). It is possible that some of the very heterogeneous telomere length profiles are indicative of this rarer recombinational based mechanism for telomere elongation and thus apoptotic avoidance.

5.10 Conclusions and future work:

High resolution analysis of individual XpYp telomere lengths within this cohort has revealed a significant association between telomere length and survival statistics. Telomeres below the threshold length of 2.26kb have previously been shown to be dysfunctional when observing

the high frequency of fusions involving telomeres of this length and below (Capper et al. 2007). This analysis of 120 patients with invasive ductal carcinoma of the breast has further confirmed the strong prognostic power of telomere length. Further work with a larger cohort of patient samples with varying breast cancer subtypes would be required to substantiate the correlations between short telomere length and poor prognosis. An interesting aspect would be to analyse telomere lengths in samples with varying severity of prognosis in breast neoplasia. Using STELA, along with a much larger dataset in breast cancer could provide a potential powerful prognostic marker not just for breast cancer, but also for human diseases that are characterised by chromosomal instability.

Chapter 6: an Analysis of Telomere fusion in Invasive Ductal Carcinoma

6.1 Abstract

The successive shortening of human telomeres as a natural function of age can induce replicative senescence (M1), which is the arrest of cellular division in response to DNA damage via the p53 and pRb (cell-cycle checkpoint) pathways. This process allows telomeres to act as a tumour suppressor mechanism that confers a barrier to unlimited cell division. In the absence of a functional DNA damage response, cells can bypass replicative senescence and telomeres erode further until cells are in the crisis/M2 stage. Telomeres in crisis have lost their ability to form a protective cap at the telomeric ends. Fusion of short dysfunctional telomeres can result in the formation of dicentric chromosomes and the initiation of cycles of anaphase bridging, breakage and fusion, which in turn can cause large-scale, potentially oncogenic, genomic rearrangements such as non-reciprocal translocations, aneuploidy and gene amplification.

To examine the extent of telomere fusion in breast cancer I applied a single molecule telomere fusion assay to 40 patients with invasive ductal carcinoma (IDC). Unfortunately the dilute nature of the DNA extractions received from the Wales Cancer Bank limited the detection and sequencing of numerous putative fusion events. However putative telomere fusion events were detected; only one of which could be characterised at the DNA sequence level this was observed in one patient with grade III IDC, and a telomere length mean of 1.79kb. This event involved the XpYp and XqYq telomeres and may have resulted in the generation of an Xp/Xq ring chromosome Whilst preliminary these findings are consistent with the view that short telomeres in IDC are capable of fusion and that this may in turn drive genomic instability in this disease.

6.2 Introduction to experiments

6.2.1 Genomic instability and telomere dysfunction in Cancer:

Genomic instability is thought to be a key mechanism by which cells accumulate a succession of mutations needed for cancer progression (Loeb, 2001; Kwei et al., 2011). There are many hallmarks of genomic instability that have been observed in different types of cancer (Lengauer et al., 1997; Negrini et al, 2010). One of which is the presence of aneuploidy. Aneuploidy is a way of describing cells that have an abnormal number of chromosomes resulting in incorrect chromosomal segregation during cell division (Sen, 2000) and has been extensively documented as a hallmark in many types of cancer (Reviewed by Ross et al., 2003). In colorectal carcinoma aneuploidy has been extensively correlated with progression to malignancy via an increase in detectable aneuploidy at the adenoma-carcinoma transition (Okamoto et al., 1988; Reid et al., 1996; Rudolph et al., 2003).

6.2.2 Genome instability and telomere dysfunction in cancer

Telomeres have been shown to provide a protective 'cap' and thus have a key role in upholding genomic integrity (Murnane, 2006, Capper et al., 2007). Several studies have correlated hallmarks of genomic instability with an increase in telomere dysfunction and have thus proposed telomere dysfunction as a mechanism that gives rise to types of genomic instability (Artandi and DePinho, 2000; Rudolph et al., 2001). Telomere dysfunction can lead to the formation of dicentric chromosomes/anaphase and this causes cycles of BFB (McClintock, 1941). Cycles of BFB can result in the loss/gain of genetic material or NRTs (Lo et al., 2002), compromising the integrity of the genome and thus forming potentially oncogenic rearrangements. *In vivo* experiments using telomerase knockout mice, showed an increase in detectable NRTs correlated with shorter telomeres, and an increase in the formation of epithelial tumours (Artandi et al., 2000). Another study by Rudolph and colleagues used anaphase bridge formation as a hallmark of telomere dysfunction, by way of

studying telomere dysfunction within a human malignancy (Rudolph et al., 2001). Sporadic human colorectal tumours at different prognostic stages were observed microscopically to determine the proportion of anaphase bridges, and found that the proportion of anaphase bridges increased with higher colorectal tumour grades (Rudolph et al., 2001). This increase in anaphase bridges was not however consistent in metastatic cells and it was thus hypothesised that telomere dysfunction is a driving force behind genomic instability but is an early mechanism driving transition from benign to malignant stages (Rudolph et al., 2001).

6.2.3 Telomere fusions detected in cancer: Telomere fusions in Osteosarcoma

Osteosarcoma (OSA) is the most common form of bone cancer found in mammals and canines and is known to display a large spectrum of karyotypic abnormalities. *FISH* was applied to canine OSA cell lines in order to study telomere dysfunction via the presence of telomere fusions as a novel prognostic marker (Maeda et al., 2012). Findings revealed numerous telomere fusions and interstitial telomere signals in all OSA cell lines that were also displaying DNA damage (assessed by co-localisation of γH2AX and telomere signals in interphase cells). Telomere fusion analysis was thus hypothesised as a possible diagnostic tool.

6.2.4 Telomere fusions in Haematological cancer

In CLL clear evidence of large scale genomic instability was observed in samples displaying telomere dysfunction (Lin et al., 2010). Telomere dysfunction was assessed by using STELA coupled with a single molecule assay to detect telomere-telomere fusions. Work by Lin et al. showed a direct observation of short telomeres fusing and higher fusion frequencies were correlated with increasing severity of CLL according to Binet's staging (Lin et al., 2010). This is a good example of how dysfunctional telomeres could be driving the progression to

malignancy. Sequence analysis of telomere fusion events from CLL patient samples showed certain mutational characteristics such as large-deletion events that included both the telomere and sub-telomeric DNA (Lin et al, 2010). Telomeres were also shown to fuse close to fragile sites and regions of microhomology at the point of telomere-telomere fusion, which may provide clues to how telomeres fuse and whether they have a tendency to fuse at certain loci (Lin et al, 2010).

6.2.5 Genomic instability and telomere dysfunction in breast cancer

Studies have proposed a link between telomere dysfunction, genomic instabilities and the progression of cells to malignancy (Rudolph et al., 2001). Demonstrations of this phenomenon of early telomere dysfunction driving genomic instability and progression of breast cancer are circumstantial. One study found an increase in anaphase bridges as a marker for telomere fusion in breast cancer samples via fluorescent *in situ* analysis (Chin et al., 2004). Human mammary epithelial cells in crisis in culture were compared with different breast cancer patient samples in different stages (ranging from atypical to invasive) to support their hypothesis/model that transition through telomere crisis is a crucial event in the progression of breast carcinomas. Observations of short telomeres and an increase in anaphase bridges were correlated with the adenoma to carcinoma transition in breast cancer samples implicating telomere dysfunction as an early event in breast cancer. These studies were carried out in a relatively small patient cohort of 31 patient samples, 4 of which displayed anaphase bridges (Chin et al., 2004). Telomere length analysis by *FISH* is limited in its ability to detect telomeres that are short enough to be in crisis due to the limited capacity of the probe to hybridise to shorter telomere repeat arrays (Baird et al., 2003). The 'short' telomeres in this cohort were defined as being 4.5 kb, however telomere thresholds of cells in crisis below which they are believed to be dysfunctional have been detected as below 1kb (Capper et al. 2007). Therefore, though there are examples of correlations between

telomeres and genomic instability in the progression of breast cancer (Artandi et al., 2000; Chin et al., 2004), there has not been a direct validation that telomeres become short enough to lose their end-capping function and undergo fusion.

6.2.6 Detecting Telomere Fusions *in vitro* using Single molecule techniques

There are limitations with a lot of current techniques of detecting telomeres short enough to warrant them being dysfunctional, i.e. that have lost enough telomere repeats to have lost their protective cap. The development by Baird and colleagues of a single molecule telomere length analysis and high resolution fusion PCR has since enabled the characterisation of fusion events both *in vitro* and *in vivo* (Capper et al., 2007; Letsolo et al., 2010; Lin et al., 2010). STELA or 'single telomere length analysis' has been used to detect short telomeres that are of a length at which they readily undergo fusions. Cells that have been forced to bypass senescence have been shown to lose telomere repeats and increasingly fuse with other telomeres (Capper et al., 2007). This telomere fusion assay was initially developed for the well characterised chromosome ends XpYp and 17p, but has since been expanded to encompass around 43% possible telomere-telomere fusions in the human genome by including oligonucleotides to detect fusions in the 16p and 21q groups of telomeres. Using the additional 16p and 21q primers in the existing assay for telomerase expressing HEK 293 cells and E6/E7 HPV expressing MRC5 cells, Letsolo et al. found an increase in the number of detectable fusion events. Furthermore an increase in the number of detectable fusion events compared with the original XpYp/17p fusion analysis was observed. All the fusion events that were sequenced involved the deletion of either one or both of the telomeres, these deletion events extended into the telomere-adjacent DNA. Microhomology was observed at multiple fusion junctions. Due to the characteristics of the sequenced fusion events, an alternative NHEJ mechanism, likely a microhomology mediated pathway (MMEJ) that is

error prone and biased towards G:C base pairs has been hypothesised as the repair process that is enabling the telomere-telomere fusions.

6.3: Project Aims

The aim of this work was to use single-molecule tools for detecting telomere fusion to examine the frequency and nature of fusion in breast cancer. These assays were applied to a panel of 36 samples of tissue DNA from invasive ductal carcinoma of the breast. This will test the hypothesis that telomere fusion occurs in breast cancer and has the potential to drive genomic instability and clinical progression.

6.4: Results: Detecting fusions in patients with short telomere distributions

A single molecule PCR-based assay was developed previously using oligonucleotide primers targeted to the telomere-adjacent sequences of specific telomeres, to around 6 kb from the telomere repeat array pointing toward the telomere (Capper et al., 2007; Letsolo 2010). Detection of specific products by Southern hybridisation with telomere specific probes allows telomere-telomere fusion events to be visualised. So far this assay has been designed for used with XpYp, 17p, 21q and 16p specific primers (Figure 6.1).

Figure 6.1: Telomere fusion oligonucleotide primer design:

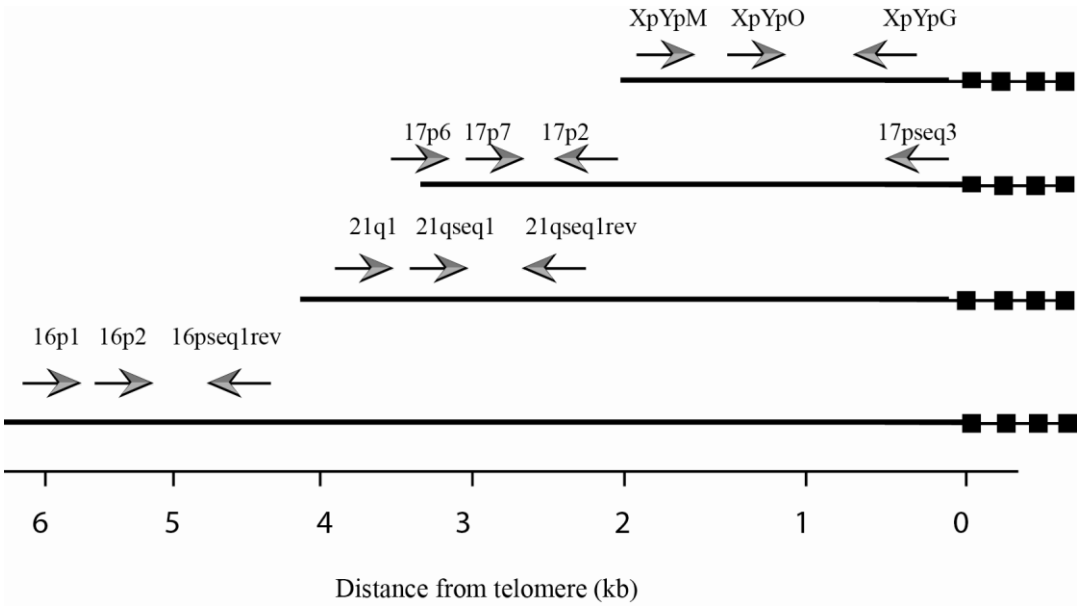


Figure 6.1: Figure shows the relative position and orientation of various oligonucleotides used in the fusion PCR and probe detection on the Southern blot. Black lines represent chromosomes and the black squares on the right hand side of each chromosome represent telomere repeat arrays. Adapted from Letsolo et al. 2010

Using the telomere fusion assays on cells undergoing crisis in culture (Capper et al., 2007 and Letsolo et al., 2010) and CLL samples *ex vivo* (Lin et al., 2010) it was apparent that cells exhibiting shorter telomere length profiles display higher frequencies of fusion events.

Breast DNA samples with the shortest mean telomere lengths were therefore selected for use in the fusion assay. Both mean telomere lengths and the telomere length profiles were taken into account when choosing patients to screen for telomere fusions. Similar experiments in CLL have revealed a threshold telomere length detectable by STELA below which telomeres become capable of fusion (3.8 kb; Lin et al. *in prep*). This meant that telomeres around or below this threshold were selected first for use in the fusion assay. MCF7 and HEK-293 cells both express telomerase and maintain their telomeres at a short length and were used as positive controls for fusion analysis studies. Figures 6.2, 6.3 and 6.4 show examples of the STELA profiles of patients with short telomere distributions and the resulting fusion bands detected by southern blot at 17p (figure 6.2), 21q (figure 6.3) and 16p (figure 6.4) telomeres.

Figure 6.2: STELA profile with associated fusion band detection at 17p

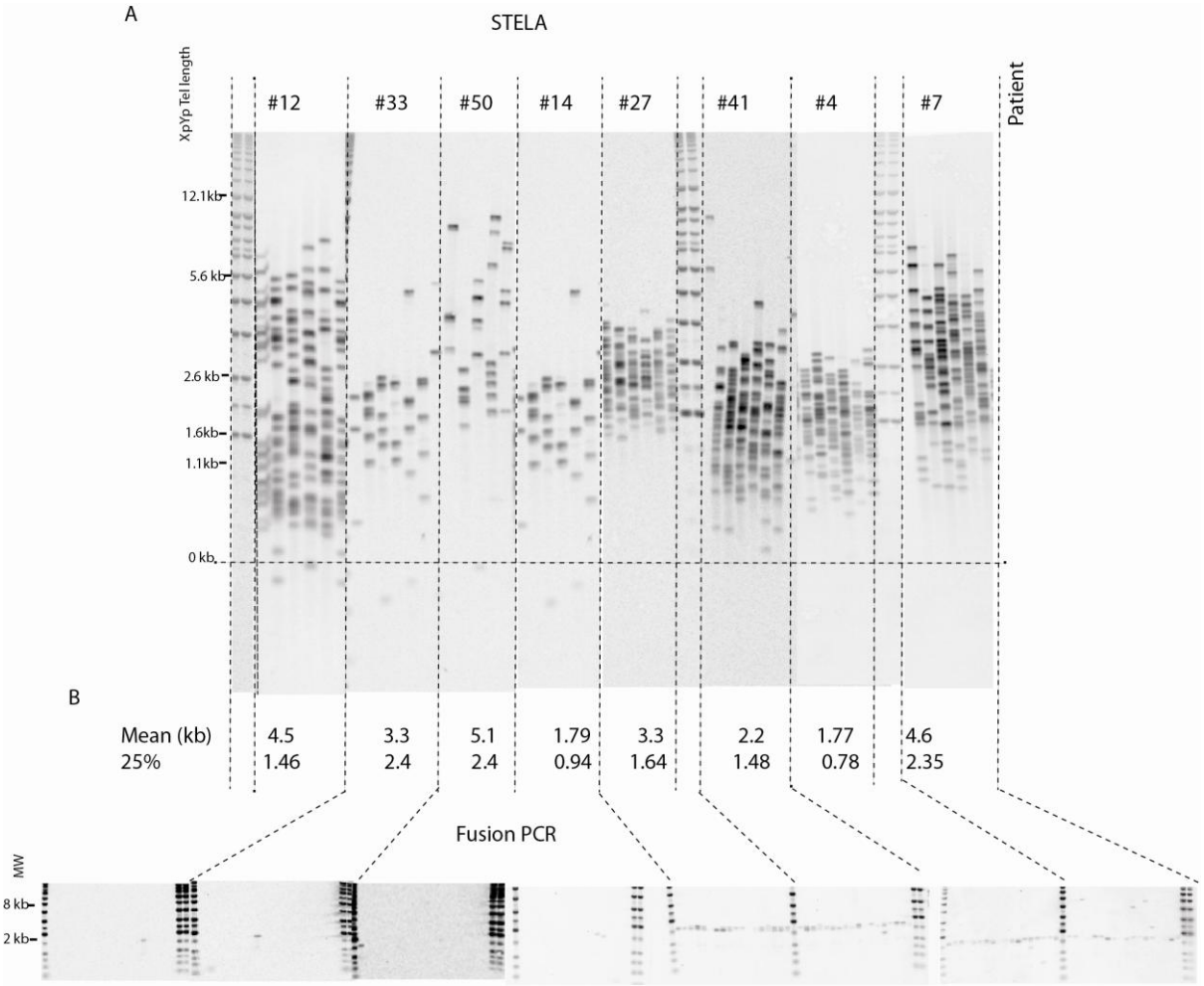


Figure 6.2: (A) STELA blot for 8 patients with IDC showing mean telomere length (kb) and lower 25th percentile below each patient. **(B)** The resulting fusion bands were detected by Southern hybridisation with the 17p telomere adjacent probe.

Figure 6.3: STELA profile with associated fusion band detection at 21q

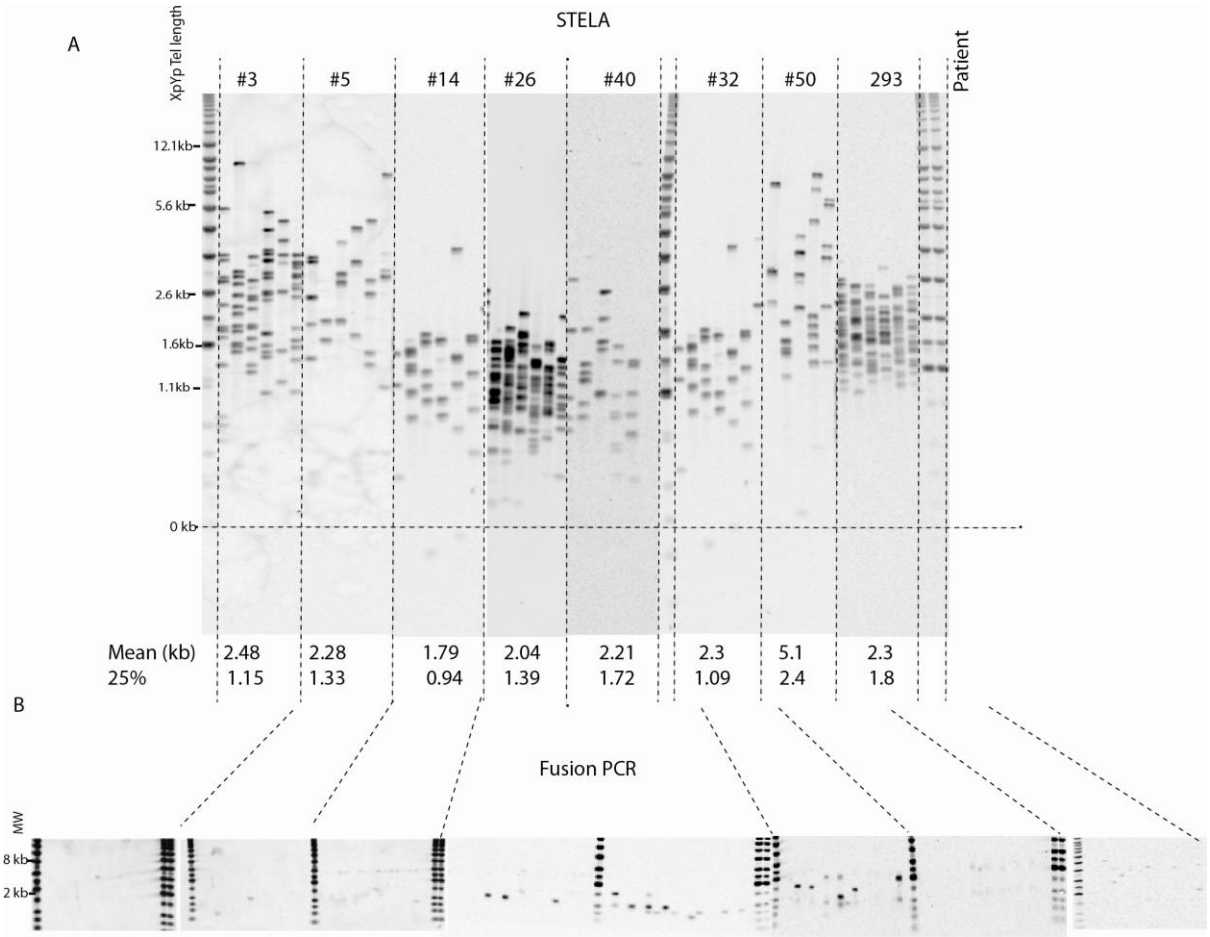


Figure 6.3: (A) STELA blot for 8 patients with IDC plus a 293 control (HEK-293 cells) showing mean telomere length (kb) and lower 25th percentile below each patient. **(B)** The resulting fusion bands were detected by Southern hybridisation with the 21q1 telomere adjacent probe

Figure 6.4: STELA profile with associated fusion band detection at 16p

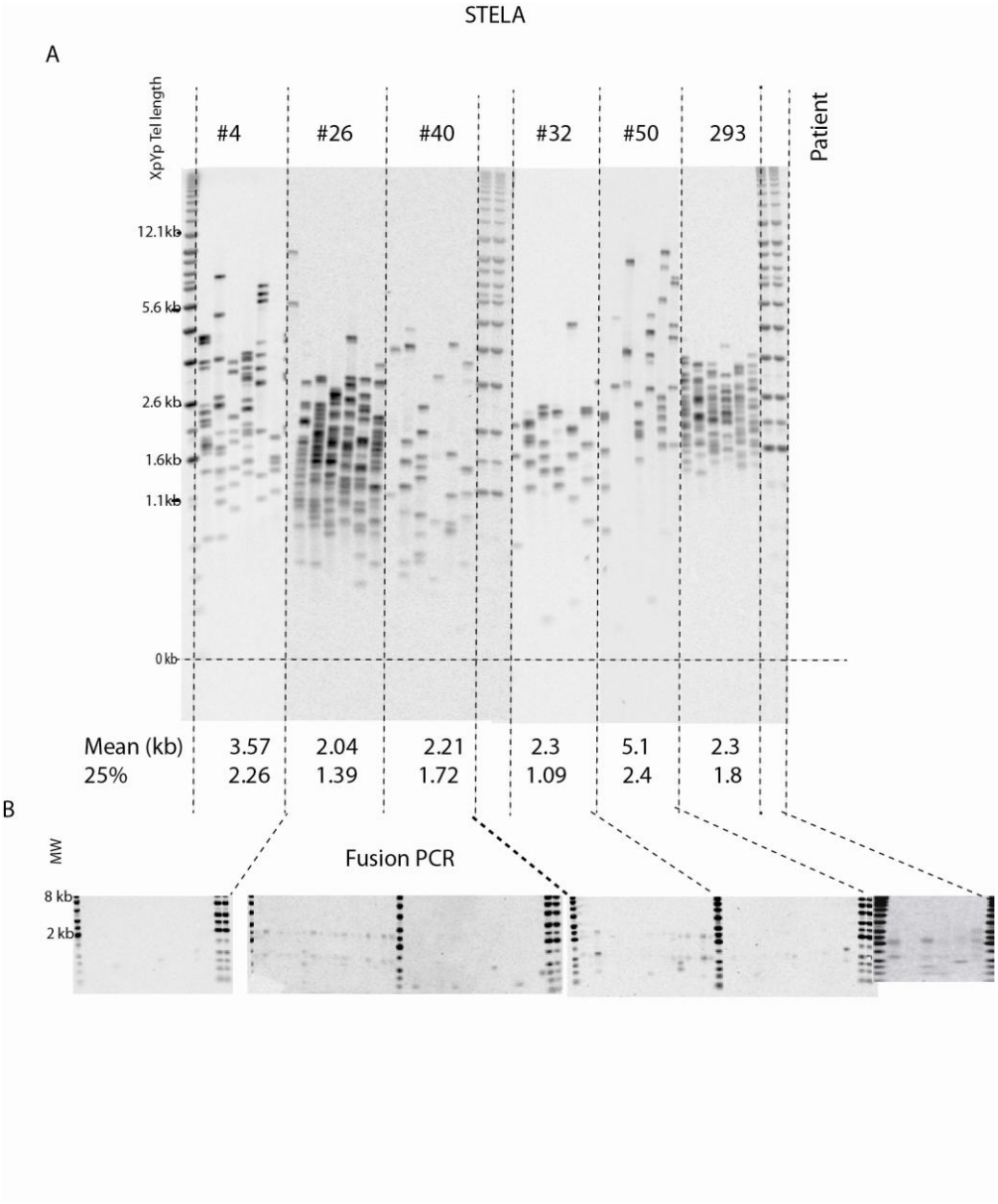


Figure 6.4: (A) STELA blot for 5 patients with IDC plus a 293 control (HEK-293 cells) showing mean telomere length (kb) and lower 25th percentile below each patient. **(B)** The resulting fusion bands were detected by Southern hybridisation with the 16p telomere adjacent probe

6.5 Constant fusion product

Hybridisation of Southern blots with 21q and 16p probes resulted in the production of a constant banding pattern with products of the same size across multiple reactions from different individuals. This type of pattern has been observed previously in our lab (Letsolo et al 2010) when using the 21q and 16p probes are thought to result from cross-hybridisation of probes with interstitial loci that share homology with the families of each telomere group. These therefore represent an internal DNA loading control and control for the Southern hybridisation. Typically 100-200ng of high molecular weight DNA are analysed per fusion reaction, this results in strong uniformly hybridising constant bands with the 21q and 16p probes (See Figure 6.5) (Letsolo et al 2010). In contrast, the fusion analysis of the breast cancer samples showed that the uniform constant band is often not present, or present only in occasional reactions in any one sample (Figure 6.6) This is consistent with dilution of this DNA molecule to the single molecule level. Unfortunately the concentrations of DNA available from the breast cancer samples provided by the Wales Cancer Bank were too low for an adequate analysis with a mean concentration of just 31ng/ul. Most of these samples were below the detectable range of the spectrophotometer or the fluorometer used to estimate the concentration of DNA. Thus the analysis of the breast cancer sample was suboptimal, but nevertheless some useful data could be obtained.

Figure 6.5: Typical Fusion Reactions using 100 ng HEK-293 DNA

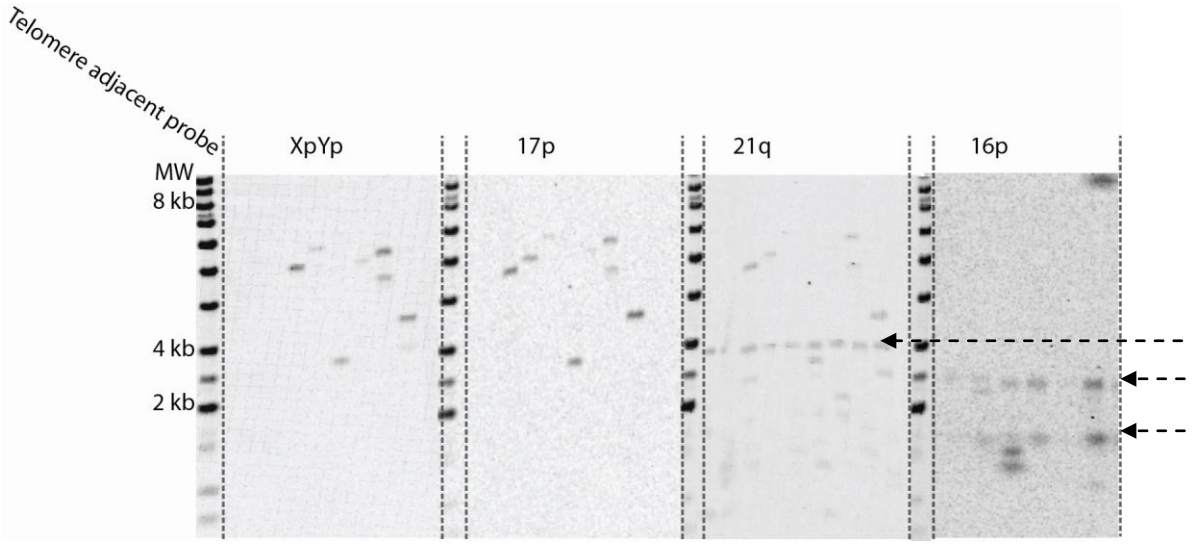


Figure 6.5: The outcome of a standard fusion PCR assay with sufficient concentrations (100ng/ul) 293 DNA. On hybridisations with 21q and 16p telomere adjacent probes, a constant band is detected indicated in figure by arrows on right hand side.

Figure 6.6: Constant band in IDC patient samples compared with 293 controls

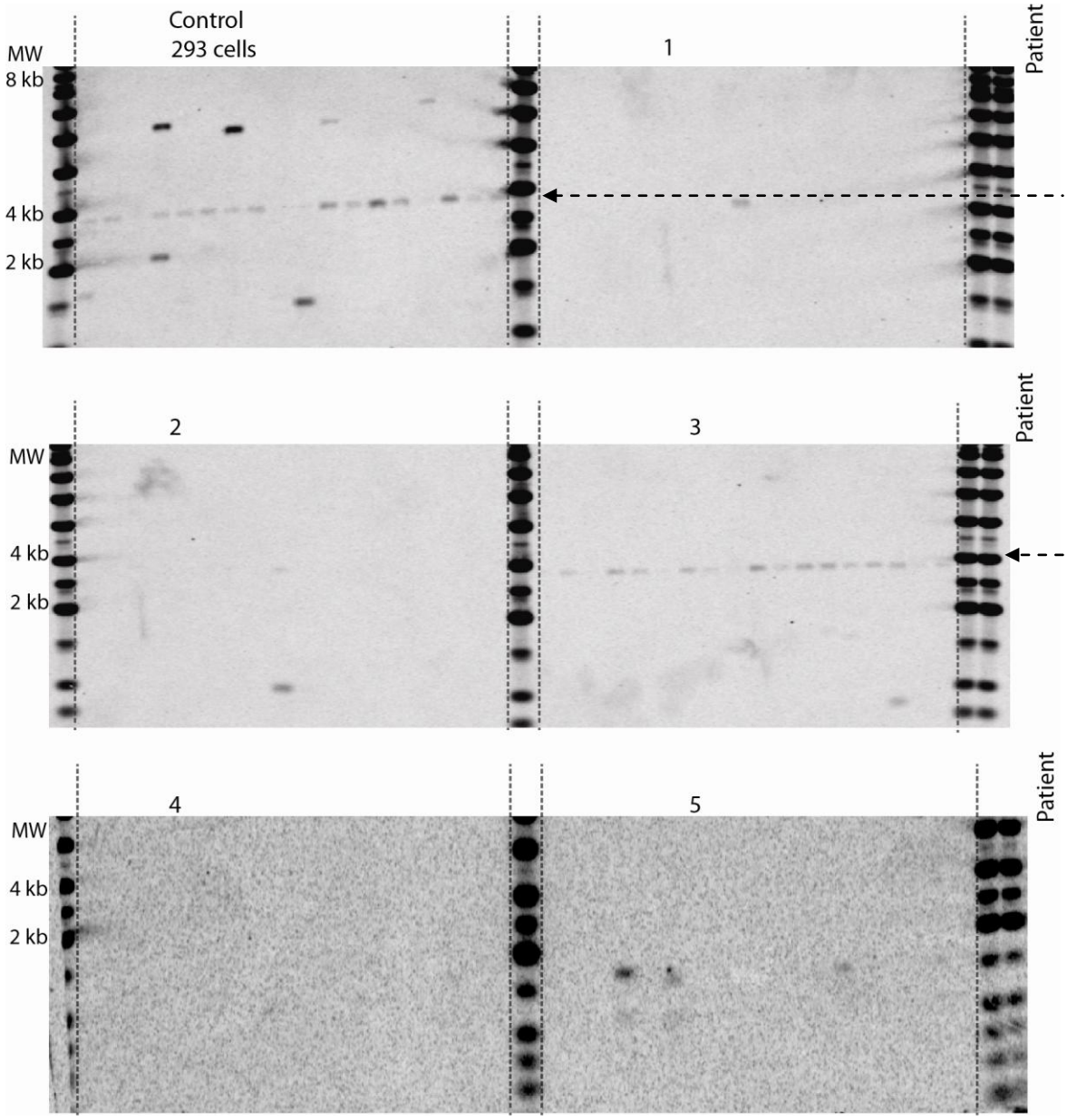


Figure 6.6: Constant band detection was highly variable in IDC patient samples when compared with 293 controls. Some patients had no bands at all (patient #4) and some had sporadic bands that appeared to be constant in size (#5). Constant banding patterns are highlighted in figure by arrows on right hand side.

6.6 Fusion frequencies of detected bands by Southern blot

Fusion frequencies can be calculated based on the number of input molecules. Input molecules can be estimated by using the approximate molecular weight of a single diploid human genome, which is 6 picograms and this can be related to the number of detected amplified molecules. Per fusion reaction the diploid genome equivalents are 5001 (30ng) (166.7 molecules/ng). For each patient 18 reactions were carried out representing a total of 90000 input genome equivalents, on this basis fusion frequencies were calculated. Due to the dilute nature of the DNA (average total concentration of 31ng/ul; ranging from 9ng/ul to 67ng/ul) the concentrations used in each fusion reaction is an estimation (DNA quantification is less accurate with dilute DNA), thus some samples may have had less input DNA than others and this will affect the fusion frequencies calculated. All fusion frequencies for each patient are in Appendix 4 Table C, however the averages for each telomere are shown below (Table 6.1). The majority of fusion bands visible by Southern blot involved the 21q and 16p telomere groups, however as mentioned earlier, shared homology between telomere groups and the presence of constant bands made distinguishing real fusion events difficult. Thus telomere fusion frequencies were calculated involving the XpYp and 17p telomeres. The mean fusion frequency across all samples is 3.4×10^{-5}, which was slightly lower, but still close to the with frequencies observed in some CLL patients and MRC5 cells in crisis; patients with Binet's stage C had a fusion frequency of 7×10^{-5} , and MRC5 E6/E7 (PD 46) had a frequency of 8×10^{-4} (Lin et al., 2010).

Table 6.1: Fusion band detection frequencies: Average frequencies per blot.

Experiment 1: Patients 1-9	Fusion frequency	Mean
XpYp	0	
17p	1.2×10^{-5}	
21q	2.5×10^{-5}	1.9×10^{-5}
Experiment 2 : Patients 1-8		
XpYp	0	
17p	8.8×10^{-5}	
21q	1.8×10^{-5}	
16p	6.6×10^{-5}	5.7×10^{-5}
Experiment 3: Patients 1-8		
XpYp	0	
17p	1.1×10^{-5}	
21q	3.3×10^{-5}	
16p	3.1×10^{-5}	2.5×10^{-5}
Experiment 4: Patients 1-8		
XpYp	2.2×10^{-5}	
17p	3.3×10^{-5}	
21q	8.8×10^{-5}	
16p	1.5×10^{-5}	3.9×10^{-5}
Experiment 5: Patients 1-3		
XpYp		
17p	3.3×10^{-5}	3.3×10^{-5}
TOTAL AVERAGE		3.4×10^{-5}

Table 6.1 : Average fusion frequencies calculated on a blot by blot basis

6.7 Sequencing of telomere-telomere fusion events

The DNA sequence of putative telomere fusion events detected with the fusion assay can be characterised following purification by re-amplification with nested PCR primers.

Unfortunately with the breast cancer samples the re-amplification of fusion bands was largely unsuccessful. A few bands were sequenced but results showed that the majority of bands were just artefacts from the PCR reaction. One band however was successfully sequenced and the fusion event characterised. The band was detected using both the XpYp and 16p telomere probes (Figure 6.7). Sequence analysis revealed it to be an intra-chromosomal fusion event involving the XpYp and XqYq telomeres. The band was detected using the 16p probe on the Southern blot because XqYq is part of the 16p telomere family sharing over 95% homology. Alongside this telomere fusion was a deletion of 263 bp from the XpYp telomere and a deletion of 3900 bp from the XqYq telomere. Below (Figure 6.8) shows the sequence at the fusion junction.

Figure 6.7: Southern blot showing fusion bands that were sequenced

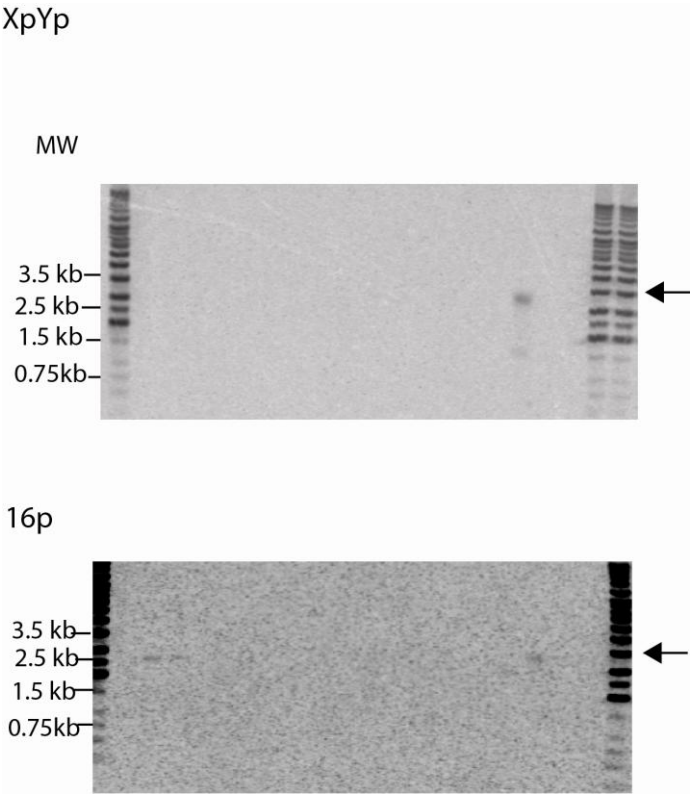


Figure 6.7: Southern blots show the result of fusion band detection for the same patient sample with two separate hybridisations, first (top) with the XpYp specific probe and second (bottom) with the 16p specific probe. A similar sized band can be on the right of each blot at about 3.5 kb (shown by arrow head).

Figure 6.8: DNA sequence at the junction pont for XpYp/XqYq telomere fusion: a putative ring Chromosome.

a.) XpYp ...GTTGTGGTCTCACTGGCTCAGGACT TGATGGCACCTCCCTCCCTCTCT...XqYq

-263 bp -3900 bp

b.)

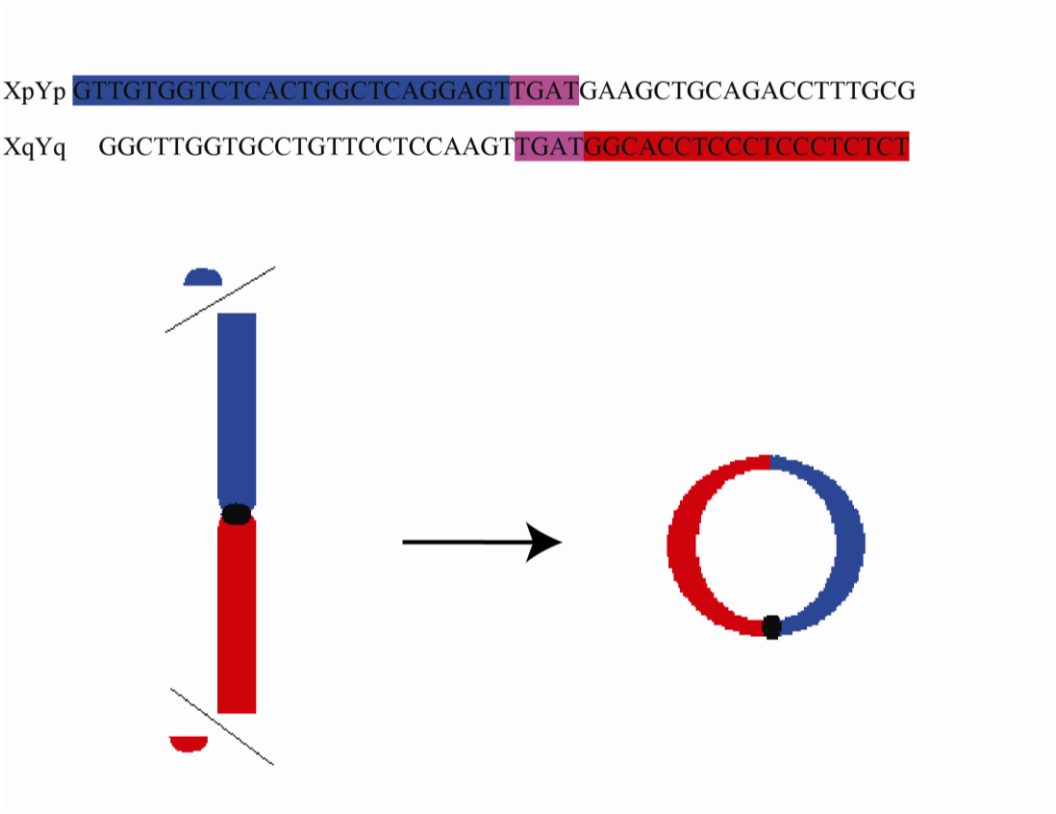


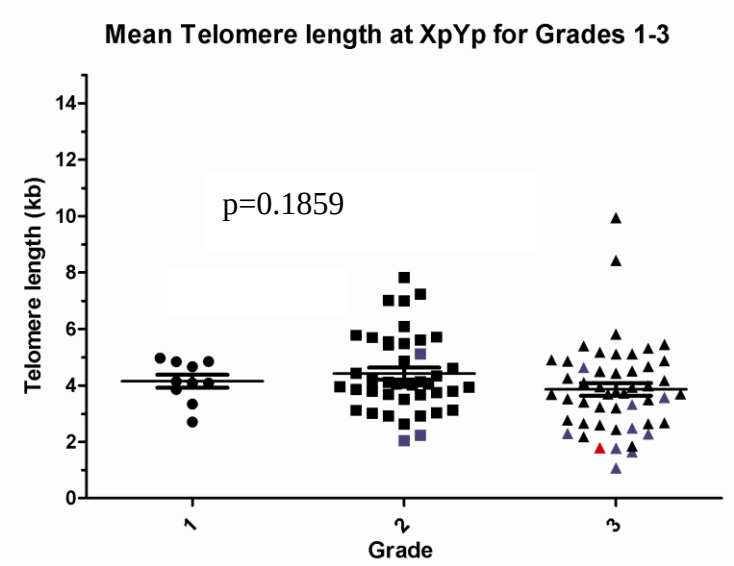
Figure 6.8: Top figure (6.8a) shows the sequence of the fusion junction sequenced following re-amplification of the bands displayed in figure 6.7. Below (6.8b) shows a graphical representation of how the putative ring chromosome may form, first by a large scale deletion then by ligation of long and short arms.

6.8 Correlating telomere fusions with grades I-III in invasive ductal carcinoma

Studies in CLL have shown a correlation between an increased fusion frequency and more progressive prognostic stages (Lin et al., 2010). It is hard to make any correlations in this way for breast cancer due to the limitation in sequencing telomere-telomere fusions. The one such fusion that was fully characterised was detected in a patient in the most severe prognostic group (grade III) with a telomere length below the threshold thought to confer telomere dysfunction (telomere length detected = 1.79kb; threshold =3.8kb (Lin et al *in prep)*). Patients displaying a high number of fusion bands were highlighted on a scatter plot to see if there was any relation to telomere length and grade (Figure 6.9). No fusions were found for the lowest grade which would be consistent with studies in CLL (Lin et al., 2010). This however

needs to be verified by testing the entire cohort and by using better quality DNA in the hope to sequence more fusion events so that conclusions can be drawn.

Figure 6.9: Summary of fusions on scatter plot of patient telomere mean lengths:



One-way analysis of variance	
P value	0.1859
P value summary	ns
Are means signif. different? (P < 0.05)	No
Number of groups	3
F	1.712
R squared	0.03342
Bartlett's test for equal variances	
Bartlett's statistic (corrected)	6.853
P value	0.0325
P value summary	*
Do the variances differ signif. (P < 0.05)	Yes

Figure 6.9: Scatter plot showing patient telomere means for all patients categorised by histological grade (NPI). Patient in red is the one that had a ring fusion. Patients in blue are those with putative fusions that were not successfully sequenced thus characterised. One way ANOVA comparing the telomere length means across three grades at XpYp showed no significant difference (p=0.1856)

6.9 Discussion

6.9.1 Limitations of detecting telomere-telomere fusions

The ability to detect telomere-telomere fusions occurring within IDC samples has been limited. The main reason for this limitation is likely to be that the starting DNA concentrations were extremely dilute ('Materials and Methods' chapter has the list of DNA concentrations for each patient sample). Fusion PCR has been shown to be most effective at detecting single molecule fusion events with starting DNA concentrations of between 100ng/ul and 200ng/ul (Capper et al., 2007). Many of the IDC samples only had 10-30ng/ul to start with and possibly less in some cases, but low concentrations are difficult to detect and quantify accurately. STELA analysis was much more feasible and works well with just 250 pg of DNA per reaction. Telomere fusion in contrast is rare and thus more input DNA molecules are required in order to provide detectable molecules. A gross underestimation of the amount of telomere fusion in IDC may mean that there could be many more telomere fusion events in these sample, as has been seen in CLL, however there is not enough DNA material to draw accurate conclusions. Array-CGH would be a useful tool to determine the extent of genomic instability for each patient (Solinas-Toldo et al., 1997; Pinkel et al., 1998), but again the lack of sufficient DNA starting concentrations would limit this.

Another explanation for the lack of fusions that have been characterised is that there are no telomere fusions present in IDC; perhaps telomere fusion in breast cancer is an early event driving genomic instability and thus occurs more frequently in the *in situ* stages as opposed to the later invasive stage being examined here. However, recent findings from another lab using similar technology suggest that telomere fusions can be detected at similar frequencies during early ductal *in situ* and in the later invasive ductal carcinoma stage (Tanaka et al., 2012). Primers targeted towards TAR-1 (telomere associated repeat-1) subtelomeric regions were used to sequence putative fusion junctions in both pre-invasive and invasive breast

cancer samples. Sequencing the fusion junctions indicate that telomere fusions are frequent in the both stages of breast cancer and could thus telomere dysfunction could be an early event in cancer progression. Secondly consistent with findings in CLL interstitial sites and large regions of microhomology were detected in the sequencing of fusion junctions (Tanaka et al., 2012; Lin et al., 2010). This may implicate a microhomology mediated mechanism of telomere fusion that has been hypothesised previously (Capper et al., 2007; Letsolo et al., 2010, Lin et al., 2010). Solid breast tumours were readily available for this recent work by Tanaka et al. and thus a higher concentration of DNA was used in experiments thus increasing the ability to detect telomere fusion events.

This PCR based fusion assay used in these experiments designed by Baird et al. has been successfully used to characterise mutational telomere fusion events in colorectal carcinoma, CLL, epidermal carcinoma and HPV to name a few, and 48% telomeres are encompassed by this assay. However not all telomeres are included in this assay and other artefacts are commonly amplified as well, especially when using the 21q and 16p telomere adjacent oligonucleotide probes. Only two fusions were detected for XpYp, which is the most unique telomere in terms of sequence homology with other telomere families so would be the most accurate. There is a SINE and a minisatellite repeat array in proximity to the XpYp telomere (Baird et al., 1999) and the presence of such satellites adjacent to the telomere could limit the detection of fusion events occurring in those regions; repeat DNA might therefore be refractory to the fusion PCR reaction.

6.9.2 Conclusions and future directions

One telomere fusion event has been successfully characterised in IDC, and interestingly this

fusion event was potentially a rare ring chromosome and also occurred in a patient with one of the shortest mean telomere length distributions (1.79 kb). Even though there was a limitation in the number of fusions that could be found in the IDC it has been hypothesised that the main reason for this was the starting DNA concentrations, as similar studies in other types of cancer such as CLL and more recently using a similar technique in the same type of breast cancer has revealed that short telomeres exhibit a high frequency of telomere fusion. Telomere fusion could thus be used as a marker for genomic instability in the progression of cancer, however for these experiments a larger dataset with the availability of original tumour tissue would be required to draw more accurate conclusions on the presence of telomere fusions in breast cancer. It would also be interesting to assess telomere fusion in a wider range of breast cancer subtypes and varying severity from pre-invasive to metastatic to observe any correlation between prognosis and telomere fusion frequency as has been observed in CLL.

Chapter 7: General Discussion

7.1 Summary: Telomere length as a biomarker in breast cancer

This project aimed to test the hypothesis that telomere dysfunction occurs during the progression of breast cancer and that this can drive the large-scale genomic rearrangement frequently observed in this disease.

Short telomeres that are below lengths at which they are able to function are thought to drive genomic instabilities such as NRTs and cycles of anaphase B-F-B, from which oncogenic rearrangements can arise (Baird et al., 2006, Lin et al., 2010). Studies in CLL have shown that telomeres length below a threshold of 2.26kb are capable of undergoing a high frequency of telomere-telomere fusions (Britt-Compton et al., 2006; Capper et al, 2007; Lin et al., 2010; Lin et al. in prep) and that the presence of telomeres below the threshold was correlated with poor prognosis. The same techniques were thus applied to a cohort of invasive breast cancer samples (IDC) to assess whether telomere length could be related to prognosis in this malignancy. Applying a single molecule long range PCR strategy (STELA) enabled the characterisation of all individual telomere lengths within each patient sample with a high degree of accuracy for telomeres XpYp and 17p, but not at 2p. STELA analysis of these samples provided evidence of the shortest telomere lengths detected in a breast cancer setting; these lengths that are below the detectable range of the other telomere-length analysis techniques.

7.2 The importance of biomarker detection

Biomarker detection is becoming increasingly important in breast cancer for patient stratification with a view to providing more targeted treatment options that may improve patient survival and also quality of life. Due to the heterogenic nature of breast cancer, it is difficult to separate out individual patients for different therapies, and as a result as many as

80% of patients receive some form of adjuvant therapy. It is disputed as to whether this 'over-treating' with cytotoxic chemicals is necessary (Weigelt et al., 2005). To date, a handful of studies have studied the link between telomere length and breast cancer risk, however these have not proved powerful enough to be translated into clinical practice.

Oxidative damage, a cause of telomere shortening and hence a major cause of cellular ageing, has been shown to increase genomic instability in carcinomas. Reactive oxygen species (ROS) cause damage to DNA, producing mutations. A tumoural increase in hydrogen peroxide (reactive oxygen intermediate) levels has been linked with an increase in genomic instability and therefore an increase in malignant and metastatic potential of tumours (Szatrowski et al, 1991; Brown et al, 2001). Studies using HPLC to detect a marker for ROS - 8-hydroxydeoxyguanosine (8-OHdG) - found that this marker is more prevalent in invasive carcinoma cells (Toyokuni et al, 1994). There are multiple mechanisms that could be the cause of this increase in oxidative damage in carcinomas, an obvious source would be the inadequate tumour vascular network caused by rapid angiogenesis of tumour cells. Tumour angiogenesis causes cycles of hypoxia and reperfusion (Brown et al, 2001). In breast cancer specifically thymidine phosphorylase overexpression is a common event and has been shown to generate ROS via the metabolism of thymidine (Brown et al, 2000). The metabolism of oestrogen hormones has been shown to potentially be a further source of oxidative stress (Sipe et al, 1994). One key factor in the contribution to oxidative stress in breast carcinomas is in anticancer therapies. Doxorubicin, an anthracycline used in chemotherapy, has been shown to generate superoxides (a causative of oxidative damage) (Yokomizo et al, 1995) among other chemotherapeutic chemicals. *In vitro* experiments have shown that tamoxifen therapy has the potential to induce oxidative stress in oestrogen receptor-negative human cancer cell lines (Ferlini et al., 1999). Treatment with some chemotherapies and hormonal therapies could hence be increasing the metastatic potential of tumours via oxidative stress.

Resistance to previously effective therapies such as tamoxifen for ER positive breast cancer is becoming a common problem for treatment of these subtypes (Clarke et al, 2001). Some studies have shown crosstalk between the breast cancer subtypes such as ER positive and ERBB2 positive, which can cause difficulties in knowing which therapies will be most effective (Hurtado et al., 2008). Tamoxifen resistant strains have shown an increase in ERBB2 levels (Osborne et al., 2003).

These findings not only highlight another mechanism driving genomic instability but a key reason to elucidate more accurate prognostic biomarkers.

7.3 Clinical significance of telomere lengths

Kaplan-Meier analysis was a useful tool in correlating telomere length with patient survival, and demonstrating that STELA combined with the fusion threshold provided a considerable improvement in prognostic power. Typically biomarkers such as HER2/neu and the presence of hormone receptors analysed using Kaplan plots give hazard ratios for overall survival of between 1 and 2 (Hilsenbeck *et al.* 1998); this was consistent with the data presented in this thesis where HRs for overall survival were 1.5 for HER2 (P=0.4513), 2.798 (p=0.0364) for PGR and 3.1 (p=0.02) for ER status. In contrast, high-resolution telomere length analysis in the same cohort revealed hazard ratios of up to 87086 ($p<0.0001$). Telomere lengths were tested as a means of determining hazard ratio and were stratified based on the threshold telomere length at which fusion occurs (2.26kb). The specificity of this threshold was confirmed by recursive partitioning where the highest HR was observed at the 2.26 kb threshold. Indeed even using the non-optimal telomere lengths to stratify patients, higher hazard ratios than have been calculated for other oncogenic factors/biomarkers were still observed.

Telomere erosion has been implicated to give rise to dysfunction that when DNA response checkpoints fail to repair/remove the dysfunction, in the long run can lead to chromosomal

rearrangements of the types observed in cancerous tumours (Counter et al, 1992; Capper et al, 2007, Lin et al., 2010). Telomere dysfunction has been observed in age related disease and modelled rather well in mouse models, with successive telomerase knockout showing deficiency in intrinsic telomere length barriers that could underlie the tissue deterioration observed in age related disease (Artandi et al., 2000). The threshold telomere length at which fusion occurs (2.26kb) has been used as a prognostic threshold in CLL and has been found to be highly significant ($P<0.0001$; HR=100.8 (32.1-316.4)) and has hazard ratios that are particularly high in patients with early stage disease ($P<0.0001$; HR=1353 (205-8902)) in 200 patients (Lin et al., 2013-pending publication). It is thus interesting that the exact same threshold length has been found to be so significant in an unrelated type of cancer. This finding could not only represent a powerful biomarker by which to predict survival outcome but a biological constant by which genomic instability could be defined. As has been mentioned previously 'short' telomere lengths have been documented as being a contributing factor in a wide range of pathological conditions, not only cancer but age-related diseases and premature ageing syndromes [Reviewed by Blasco, 2005]. It may be pertinent to consider whether this dysfunctional range that has found to be highly prognostic in two separate cancer types, might be applicable to other human diseases. It is unknown why this particular length (2.26kb) is significant in terms of survival. It has been observed that below threshold telomere lengths, telomeres are in a 'dysfunctional' range and are subject to a wide range of fusion events of types that are believed to occur via error prone (MMEJ) mechanisms (Capper et al., 2007, Lin et al., 2010, Letsolo et al., 2010). This increase in telomere-telomere fusions coupled with an increasing genomic instability (assessed by array-CGH) and increasing severity of prognosis all point to telomere length as an indicator of dysfunction that may be driving progression to malignancy (Lin et al., 2010, Jones et al., 2012). Thus telomeres below this dysfunctional threshold must pose a risk for an increase in genomic

instability and thus a higher risk in terms of disease outcome. This threshold length for use as a biomarker needs to be verified using a much larger cohort with longer follow up data. This cohort was relatively small (n=120), with a follow up period of 5 years. There is still a gap in the knowledge surrounding the factors that could account for relapse in patients 5 years post diagnosis (Bosco et al., 2009); several studies have indicated that recurrent breast cancer that occurs after the 5 year mark is more aggressive in nature (Brewster et al., 2008; Bosco et al., 2009). The possibility that telomere erosion is an early event in breast cancer was mentioned in chapter 5 (section 5.95), thus it would also be vital to assess telomere dynamics in a range of different breast cancer subtypes, not just by NPI grade but by histological classifications such as pre invasive stages (DCIS/LCIS), with a long follow up period, in order to assess the true prognostic value of telomere length dynamics.

In a large-scale study using NGS technology was used to compare copy number and gene expression in parallel in 2,000 breast tumours (Curtis et al., 2012). The large cohort size used and the ability to detect new sub-groups in breast cancer bringing the number from five up to at least ten makes this study powerful for the breast cancer setting (Curtis et al., 2012). This technology enables the understanding of how copy number alterations affect gene expression profiles which could help to understand how different breast cancer types respond differently to therapies. Dysfunctional telomeres that undergo fusion events could explain the genomic alterations observed in breast cancer such as copy number alterations. Another deep-sequencing study on triple-negative breast cancer demonstrated the heterogeneity of breast cancer as a disease and that no two tumours were the same (Shah et al., 2012). Defining better prognostic subgroups for this complex disease is thus becoming more important.

Knowledge concerning gene expression profiles found using deep sequencing techniques

coupled with the utilisation of telomere length as a biomarker for prognosis could be a powerful way of both diagnosing and treating breast cancer patients much more effectively, with less overtreatment with cytotoxic chemicals.

7.4 Telomere length heterogeneity in Breast Cancer

Analysis of individual telomere lengths by STELA revealed a large heterogeneity in telomere length. Heterogeneity of telomere length seen in some patient samples could be down to inter-individual variation pre determined at the zygote (Graakjaer et al., 2006, Baird et al., 2003). Alternatively the origin of this heterogeneity could be due to inter-allelic differences in telomere lengths. One way to verify this form of heterogeneity would be to apply an allele-specific STELA PCR to separate out telomere lengths in individuals containing heterozygosities in the XpYp telomere-adjacent DNA, as has been done previously (Baird et al., 2006). Some of the poorer prognosis patients, as characterised by Nottingham grading and morbidity, might be expected to exhibit clonal telomere length distributions reminiscent of clonal expansion. However subsets of patients in the poorest prognoses group (grade III) displayed a very heterogeneous pattern of telomere length and thus no pattern between grade and telomere profile could be concluded. The origins of this telomere heterogeneity could be the existence of mosiacism in the expression of genetic markers, such as the bimodal expression of oestrogen receptors (Schnitt et al., 2006). The quality of the patient samples being analysed could also explain the more heterogeneous/bimodal clusters of telomere distributions. DNA extraction was performed in a separate laboratory prior to delivery of samples and therefore the purity of these tissue samples cannot be verified. Infiltrating epithelial cells or lymphocytes with longer telomeres could be represented in the DNA samples. Purification of breast tumour specific cells prior to DNA extraction would be necessary in future experiments to ensure that breast tumour DNA alone is being analysed for

telomere length.

A subset of cancers with varying receptor status including HER2 overexpression subjected to FISH combined with PML protein immunofluorescence allowed the identification of the ALT phenotype in 3 out of 21 HER-2 positive cases and these showed a high proliferative rate (Subhawong et al., 2010). ALT is usually rare in carcinoma cell types. Phenotypically ALT telomeres have been shown to be long and heterogeneous, which could fit a few of the STELA profiles seen for this dataset of invasive ductal carcinoma, however this is hard to characterise by eye and therefore testing of patient samples for the presence of ALT markers and this could elucidate telomere maintenance mechanisms at play to explain the heterogeneous distributions.

7.5 Longer 17p telomere lengths compared with XpYp telomere lengths

Telomere lengths analysis using STELA revealed that 17p telomeres were longer than XpYp when comparing using linear regression analysis ($P<0.0001$). Is has been documented previously that the 17p telomere has a propensity to be shorter than other chromosome ends (Martens et al., 1998, Lansdorp et al., 1998) and this is believed to be as a consequence of the presence of the tumour suppressor p53 on the 17p chromosome arm. This is in contradiction with studies that have observed that 17p is not always the shortest telomere, and that some telomeres are randomly shorter (Baird et al., 2006) suggesting telomere length is governed at the zygote stage and variability could be due to difference in the maternal and paternal germlines. As mentioned earlier (Chapter 4) Telomere Variant Repeats (TVRs) at 17p may be the reason that the telomeres appear to be longer. A way of determining this in future would be to perform a TVR-PCR and deduct this from the STELA lengths. A recent study in CML (Chronic Myelogenous Leukaemia) reported longer 17p telomere lengths in patients with CML when compared with healthy patient telomere lengths using Q-*FISH* (Samassekou et al., 2011). Longer 17p telomeres could have arisen as a protective mechanism because of the

presence of the p53 gene, but this difference between 17p and other telomeres would have to be concluded with a larger dataset for a better and more conclusive comparison as to whether 17p differences are due to mechanisms important for cancer biology or whether a natural polymorphism occurrence. The development of STELA technology to encompass a larger subset of telomeres would enable a more thorough comparison of differences in chromosome arm length. This would be constructive because LOH at regions of interest containing certain tumour suppressor genes could enable further definition of mutational processes at play in breast cancer. For example some studies have shown that the loss of 9p (LOH) occurs in breast cancer due to the presence of p16/CDKN2 (Brenner and Aldaz, 1995). This could also be further verified by array-CGH.

7.6 Developing telomere fusion technology

This project also examined the hypothesis that specific regions of the genome are prone to fusion with dysfunctional telomeres. In doing so this work hoped to lead to the development of more sensitive assays for telomere dysfunction. Ligation mediated techniques aimed to target telomeres that had fused close to known restriction sites in the genome. A range of controls were successfully set up to show that the PCR based techniques would work with oligonucleotide primers designed within telomere adjacent DNA regions. Testing of these techniques revealed that their ability to detect fusions was limited. Several methods for dectecting fusions were tested and the most successful used a linker termed a ‘splinkerette’, however this technique was fairly time consuming which partly limited the characterisation of any fusion events. The results generated from the splinkerette assay were extremely dependent on the successful ligation of the splinkerette itself, a number of negative results

were generated in preliminary experiments before the ligation was successful. More fine tuning of this technique is required, however it cannot be ruled out as a method of finding telomere fusions in the genome because such techniques used in other applications are readily able to detect inserts on a smaller scales, such to isolate insertion sites in murine leukemia virus (Uren et al., 2009). The main drawback with this method was the inability to sequence and thus characterise the putative fusion bands that were visible on the southern blot. I believe that this adaptor-ligation technique was working at the single molecule level; it would therefore be worth trying to find out exactly what the banding patterns were before ruling out this technique. Sequencing the PCR products using a next generation sequencing (NGS) approach could give a clearer picture of what is being amplified. A similar technique has been used for screening HIV integrating sites using a cassette-PCR genome walking coupled with pyrosequencing (Wang et al. 2007) as a high throughput way of sequencing unknown region flanking restriction sites.

7.7 Telomere dysfunction and breast cancer

Another project aim was to assess the extent of telomere dysfunction in breast cancer. It is thought that dysfunctional telomeres are prone to telomere-telomere fusions. A telomere fusion assay was therefore applied to the breast cancer cohort used for STELA.

When E6/E7 expressing cells are forced to divide through crisis *in vitro*, erosion of telomeres has been shown to be correlated with a high fusion frequency (Capper et al., 2007). The shortest telomeres have been shown to exhibit the highest frequencies of telomere fusions in cell cultures and in CLL (Capper et al., 2007; Lin et al, 2010). Short telomeres that have lost their function are hypothesised to be a hallmark of genome instability, as a driving force behind the progression to malignancy (Baird et al., 2006, Capper et al, 2007; Lin et al., 2010). It was thus hypothesised that the patients with the shortest telomeres and the poorest clinical outcomes would exhibit a high frequency of telomere fusions as a marker of telomere

dysfunction driving breast cancer. This was not the case. Unfortunately due to the dilute nature of the breast tissue DNA, fusion analysis was limited in its ability to detect and sequence the full spectrum of fusion events that could be present. However, regardless of low starting DNA concentration, a large variety of bands reminiscent of fusion bands were visible using Southern blot analysis. These bands could be representative of fusions but the inability to sequence these bands meant that these could not be formally verified and characterised. One fusion was characterised and sequenced from a patient with short telomeres and poor prognosis. Being just one event makes it difficult to make correlations pertaining to telomere length and dysfunction; however the fact that it was present in a patient with poor clinical outcome and short telomeres is consistent with findings of a similar nature in melanoma and CLL (Letsolo, 2011; Lin et al, 2010).

7.8 Ring chromosomes

The only fusion event that was successfully sequenced in the IDC patient sample was consistent with a fusion event involving XpYp and XqYq telomeric sequences that may have resulted in a ring chromosome. This is surprising as ring chromosomes are not commonly detected in carcinomas (Gisselsson, 2002) but are more often found in sarcomas. XpYp:XqYq fusions have been frequently detected *in vitro* (Letsolo, 2011). Ring chromosomes can be formed by the loss of both chromosome arms so that they fuse together (Miller and Thurman, 2001). They can also arise due to loss of just one end causing it to stick to the opposing telomere (Arnedo et al., 2005). Loss of genetic information is not necessary for the formation of ring chromosomes. Ring chromosomes have been observed in many different diseases, such as mental retardation (ring chromome 14) (Schmidt et al., 1981), and are linked to genomic instability. Ring chromosomes are unstable structures, particularly during mitosis (Miller and Thurman, 2001; Sumner, 2003) and can form anaphase bridges. The unstable nature of ring chromosomes has implicated them in causing tumoural heterogeneity

(Gisselsson et al., 2000). Ring chromosomes are most commonly detected using hybridisation based techniques such as FISH (Bauman et al., 1980), however this can be limiting because it requires specific probes to target certain areas and metaphase spreads. FISH can also be labour intensive [Reviewed by Scouarnec and Gribble, 2011]. Sometimes a combination of SNP-a analysis and FISH is used, for example in studies in acute myeloid leukaemia (AML) a whole genome SNP-a was used to characterise a ring chromosome (Huh et al., 2012). FISH techniques have been improved to increase sensitivity using shorter probes and extended chromatin fibres termed 'FibreFISH' (Heng et al., 1992; Wiegant et al., 1992; Parra and Windle, 1993). The development of Array-CGH (Solinas-Toldo et al., 1997; Pinkel et al., 1998) and SNP arrays ((Kennedy et al., 2003; Gunderson et al., 2005; LaFramboise, 2009) have facilitated the mapping of chromosomal alterations and copy number variations. However chromosomal rearrangements can now be analysed using second generation technologies such as next generation sequencing (NGS). NGS allows sequencing in parallel of millions of DNA molecules, and more specifically the 'read-pair method' is used to study chromosome rearrangements using a paired-end read mapping technique (Tuzun et al., 2005; Korbel et al., 2007).

7.9 Conclusions and future directions:

The use of a single molecule high resolution technology (STELA) has allowed the detection of individual telomere lengths at multiple chromosome ends in patients with IDC showing a spectrum of lengths ranging from above 20kb to below 1kb. The shortest telomere length mean detected was 1.29kb, which would be undetectable using other hybridisation based techniques. Future experiments would require purification of breast tissue prior to STELA experiments to ensure an accurate quantification of telomere lengths for tumour DNA only.

A larger cohort with better DNA quality/availability would allow confirmation of these

results. A range of different breast cancer subtypes with varying severity would be necessary to compare telomere lengths among different subtypes. With a larger quantity of DNA more detailed telomere fusion analysis in different breast cancer subtypes should be carried out to screen for telomere dysfunction in breast cancer and to confirm the hypothesis further that it is this telomere dysfunction that is driving the progression to malignancy as has been correlated in CLL (Lin et al., 2010). Array-CGH or whole genome sequencing would also be possible with a better breast tissue DNA availability. This would be used to screen individuals displaying evidence of telomere dysfunction to detect any genomic rearrangements occurring and to see whether these are located primarily in telomere regions.

This project has successfully demonstrated the use STELA combined with statistical tests to show that short telomere length is highly significant in terms of prognosis in breast cancer. Telomere length stratification could thus be used as a method of defining new breast cancer subtypes in terms of severity.

ABRAHAM, R. (2001) Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes and Development*; 15: 2177-2196

ABREU, E., ARITONOVSKA, E., REICHENBACH, P., CRISTOFARI, G., CULP, B., TERNS, R. M., LINGNER, J. & TERNS, M. P. (2010) TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. *Molecular and cellular biology,* 30**,** 2971-82.

ACHUTHAN, R., BELL, S. M., ROBERTS, P., LEEK, J. P., HORGAN, K., MARKHAM, A. F., MACLENNAN, K. A. & SPEIRS, V. (2001) Genetic events during the transformation of a tamoxifen-sensitive human breast cancer cell line into a drug-resistant clone. *Cancer genetics and cytogenetics,* 130**,** 166-72.

ADAMS, S. P., HARTMAN, T. P., LIM, K. Y., CHASE, M. W., BENNETT, M. D., LEITCH, I. J. & LEITCH, A. R. (2001) Loss and recovery of Arabidopsis-type telomere repeat sequences 5'-(TTTAGGG)(n)-3' in the evolution of a major radiation of flowering plants. *Proceedings. Biological sciences / The Royal Society,* 268**,** 1541-6.

ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K., WALTER, P. (2008). "Chapter 17". *Molecular Biology of the Cell* (5th ed.). New York: Garland Science. ISBN 978-0-8153-4111-6.

ALCORTA, D., XIONG, Y., PHELPS, D., HANNON, G., BEACH, D. & BARRETT, J. C. (1996) Involvement of the cyclin-dependent kinase inhibitor p16 (INK4a) in replicative senescence of normal human fibroblasts. *Proc Natl Acad Sci USA,* 93**,** 13742-13747.

ALLRED, D. C., BROWN, P. & MEDINA, D. (2004) The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer. *Breast cancer research : BCR,* 6**,** 240-5.

ALT, F. W., BOBOILA, C., YAN, C., WESEMANN, D. R., JANKOVIC, M., WANG, J. H., MANIS, J., NUSSENZWEIG, A., NUSSENZWEIG, M. & ALT, F. W. (2010) Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. *The Journal of Experimental Medicine,* 207**,** 417-427.

AM, D., PA, H., JD, I. & JR, M. (1991) Genetic basis for p53 overexpression in human breast cancer. *Proc Natl Acad Sci USA,* 88**,** 5006-5010.

AMBROSINI, A., PAUL, S., HU, S., RIETHMAN, H. (2007) Human subtelomeric duplicon structure and organization. *Genome Biology*.;8:R151.

ANDERSON, E. (2002) Progesterone receptors – animal models and cell signaling in breast cancer The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis. 197-201.

ANDERSON, S. A. D., L (1980) Metabolism of Okazaki Fragments during Simian Virus 40

DNA replication. *The Journal of biological chemistry,* 254**,** 11495-11504.

APARICIO, O. M., BILLINGTON, B. L. & GOTTSCHLING, D. E. (1991) Modifiers of Position Effect Are Shared between Telomeric and Silent Mating-Type Loci in S . cerevisiae. *Cell,* 66**,** 1279=1287.

ARAT, N. O. & GRIFFITH, J. D. (2013) Human Rap1 Interacts Directly with Telomeric DNA and Regulates TRF2 Localization at the Telomere * □. *Journal of Biological Chemistry,* 287**,** 41583-41594.

ARTANDI, S. E., CHANG, S., LEE, S. L., ALSON, S., GOTTLIEB, G. J., CHIN, L. & DEPINHO, R. A. (2000) Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature,* 406**,** 641-5.

AUDEBERT,M., SALLES,B. and CALSOU,P. (2008) Effect of double-strand break DNA sequence on the PARP-1 NHEJ pathway. *Biochem. Biophys. Res. Commun.,* 369, 982–988.

AUTEXIER, C. & LUE, N. F. (2006) The structure and function of telomerase reverse transcriptase. *Annual review of biochemistry,* 75**,** 493-517.

AZZALIN, C. A. L., JOACHIM (2008) Telomeres: the silence is broken. *Cell Cycle,* 7**,** 1161-1165.

AZZALIN, C. M., REICHENBACH, P., KHORIAULI, L., GIULOTTO, E. & LINGNER, J. (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science (New York, N.Y.),* 318**,** 798-801.

BADCOCK, C. & CRESPI, B. (2008) Battle of the sexes may set the brain. *Nature,* 454**,** 1054-5.

BADIE, S., ESCANDELL, J. M., BOUWMAN, P., CARLOS, A. R., THANASOULA, M., GALLARDO, M. M., SURAM, A., JACO, I., BENITEZ, J., HERBIG, U., BLASCO, M. A., JONKERS, J. & TARSOUNAS, M. (2010) BRCA2 acts as a RAD51 loader to facilitate telomere replication and capping. *Nature structural & molecular biology,* 17**,** 1461-9.

BADVE, S. A. N., HARIKRISHNA (2008) Oestrogen receptor-positive breast cancer: towards bridging histopathologic and molecular classifications. *Journal of clinical pathology*.

BAI, L. & ZHU, W.-G. (2006) p53 : Structure , Function and Therapeutic Applications. *Journal of cancer molecules,* 2**,** 141-153.

BAILEY JA, CHURCH DM, VENTURA M, ROCCHI M, EICHLER EE (2004) Analysis of segmental duplications and genome assembly in the mouse. *Genome Research* 14: 789–801.

BAIRD, D. M. (2005) New developments in telomere length analysis. *Experimental gerontology,* 40**,** 363-8.

BAIRD, D. M. (2006) Telomeres. *Experimental Gerontology,* 41**,** 1223-1227.

BAIRD, D. M. (2008) Telomere dynamics in human cells. *Biochimie,* 90**,** 116-21.

BAIRD, D. M., BRITT-COMPTON, B., ROWSON, J., AMSO, N. N., GREGORY, L. & KIPLING, D. (2006) Telomere instability in the male germline. *Human molecular genetics,* 15**,** 45-51.

BAIRD DM, CHEUNG I, SCHERTZER M, BAROSS A, ROSE AM, LANSDORP PM, (2004), Strain-specific telomere length revealed by single telomere length analysis in Caenorhabditis elegans. *Nucleic Acids Res,* Volume 32, 11;.3383-3391

BAIRD DM, COLEMAN J, ROSSER ZH, ROYLE NJ. (2000) High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: implications for telomere biology and human evolution. *Am J Hum Genet*. 2000;66(1):235-250.

BAIRD, D. M., DAVIS, T., ROWSON, J., JONES, C. J. & KIPLING, D. (2004) Normal telomere erosion rates at the single cell level in Werner syndrome fibroblast cells. *Human molecular genetics,* 13**,** 1515-24.

BAIRD, D. M., JEFFREYS, A. J. & ROYLE, N. J. (1995) Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *The EMBO journal,* 14**,** 5433-43.

BAIRD DM, KIPLING D, (2004) The extent and significance of telomere loss with age.*Ann N Y Acad Sci,* Volume 1019 pp.265-268

BAIRD, D. M., ROWSON, J., WYNFORD-THOMAS, D. & KIPLING, D. (2003) Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nature genetics,* 33**,** 203-7.

BANERJEE, S., REIS-FILHO, J., STEELE, D., ASHWORTH, A., LAKHANI, S. & SMITH, I. (2006) Basal-like breast carcinomas: clinical outcome and response to chemotherapy. *Journal of clinical pathology,* 59**,** 729-735.

BARNES, D., CAMPLEJOHN, R. (1996) P53, apoptosis, and breast cancer. *Mammary Gland Biology and Neoplasia.* Apr;1(2):163-175

BATZER, M. A. & DEININGER, P. L. (2002) Alu repeats and human genomic diversity. *Nature reviews. Genetics,* 3**,** 370-9.

BAYREUTHER, K., RODEMANN, H. P., HOMMEL, R., DITTMANN, K., ALBIEZ, M. & FRANCZ, P. I. (1988) Human skin fibroblasts in vitro differentiate along a terminal cell lineage. *Proceedings of the National Academy of Sciences of the United States of America,* 85**,** 5112-6.

BAUMANN P, CECH TR. 2000. Protection of telomeres by the Ku protein in fission yeast. *Mol. Biol. Cell* 11:3265-75

BAUMANN P, CECH TR (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*.;292(5519):1171-1175.

BEDNAREK, A. K., SAHIN, A., BRENNER, A. J., JOHNSTON, D. A. & ALDAZ, C. M. (1997) Analysis of telomerase activity levels in breast cancer: positive detection at the in situ breast carcinoma stage. *Clinical cancer research : an official journal of the American Association for Cancer Research,* 3**,** 11-6.

BENTLEY,J., DIGGLE,C.P., HARNDEN,P., KNOWLES,M.A. AND KILTIE,A.E. (2004) DNA double strand break repair in human bladder cancer is error prone and involves microhomology-associated end-joining. *Nucleic Acids Res.*, 32,

BERNSTEIN, L. (2002) Epidemiology of endocrine-related risk factors for breast cancer. *Journal of mammary gland biology and neoplasia,* 7**,** 3-15.

BIANCHI, A., SMITH, S., CHONG, L., ELIAS, P. & DE LANGE, T. (1997) TRF1 is a dimer and bends telomeric DNA. *The EMBO journal,* 16**,** 1785-94.

BIANCHI A, STANSEL RM, FAIRALL L, GRIFFITH JD, RHODES D, DE LANGE T. (1999). TRF1 binds a bipartite telomeric site with extreme spatial flexibility. *EMBO J.* 18:5735–44

BILAUD, T. B., C; ANCELIN, K; KOERING, C; LAROCHE, T AND GILSON, E (1997) Telomeric localization of TRF2, a novel human telobox protein. *Nature Publishing Group,* 17**,** 236-239.

BINET, J., AUQUIER, A., DIGHIERO, G. (1981) A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer.*;48: 198-206.

BISWAS, R., VONDERHAAR, B. K., BISWAS, R. & VONDERHAAR, B. K. (1987) Role of Serum in the Prolactin Responsiveness of MCF-7 Human Breast Cancer Cells in Long-Term Tissue Culture Role of Serum in the Prolactin Responsiveness of MCF-7 Human Breast Cancer Cells in Long-Term Tissue Culture. *Cancer research,* 47**,** 3509-3514.

BLACKBURN, E. H. (1992) Telomerases. *Annual review of biochemistry,* 61**,** 113-129.

BLACKBURN, E.H. (2005), Telomerase and Cancer, *Molecular Cancer Research* ;3:477-482.

BLACKBURN, E. H. & CHALLONER, P. B. (1984) Identification of a telomeric DNA sequence in Trypanosoma brucei. *Cell,* 36**,** 447-57.

BLACKBURN, E.H. & GALL, J.G. (1978) A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol*. 120, 33–53

BLACKBURN, E.H. & SZOSTACK, J.W. (1984).The molecular structure of centromeres and telomeres. *Annu. Rev. Biochem*. 53, 163–194

BLASCO, M. A. (1997) Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* 91, 25–34 .

BLOOM, H. J. & RICHARDSON, W. W. (1957) Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *British journal of cancer,* 11**,** 359-77.

BODNAR, A. G., OUELLETTE, M., FROLKIS, M., HOLT, S. E., CHIU, C.-P., MORIN, G. B., HARLEY, C. B., SHAY, J. W., LICHTSTEINER, S. & WRIGHT, W. E. (1998) Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. *Science*, 279.
.

BODVARSDOTTIR, S. K., STEINARSDOTTIR, M., BJARNASON, H. & EYFJORD, J. E. (2012) Mutation Research / Fundamental and Molecular Mechanisms of Mutagenesis Dysfunctional telomeres in human BRCA2 mutated breast tumors and cell lines. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis,* 729**,** 90-99.

BOGAERTS J, CARDOSO F, BUYSE M, (2006). Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol*;3:540-51.

BOLDOG, F., GEMMILL, R.M., WEST, J., ROBINSON, M., ROBINSON, L., LI, E., ROCHE, J., TODD, S., WAGGONER, B., LUNDSTROM, R. (1997). Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B. *Hum. Mol. Genet*. 6: 193–203

BOJOVIC, B. & CROWE, D. L. (2013) Dysfunctional telomeres promote genomic instability and metastasis in the absence of telomerase activity in oncogene induced mammary cancer. *Molecular carcinogenesis,* 52**,** 103-17.

BOULTON SJ, JACKSON SP. (1996) Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA doublestrand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.*; 15(18):5093-5103.

BOWCOCK, A. (1993) Molecular cloning of BRCA1: a gene for early onset familial breast and ovarian cancer. *Breast Cancer Research and Treatment* Nov;28(2):121-35.

BRADLEY, M., HSU, I., HARRIS, C. (1979) Relationship between sister chromatid exchange and mutagenicity, toxicity and DNA damage. . *Nature* Nov 15;282(5736):318-20.

BRITT-COMPTON, B. & BAIRD, D. M. (2006) Intra-allelic mutation at human telomeres. *Biochemical Society transactions,* 34**,** 581-2.

BRITT-COMPTON, B., CAPPER, R., ROWSON, J. & BAIRD, D. M. (2009) Short telomeres are preferentially elongated by telomerase in human cells. *FEBS letters,* 583**,** 3076-80.

BRITT-COMPTON, B., LIN, T. T., AHMED, G., WESTON, V., JONES, R. E., FEGAN, C., OSCIER, D. G., STANKOVIC, T., PEPPER, C. & BAIRD, D. M. (2012) Extreme telomere erosion in ATM-mutated and 11q-deleted CLL patients is independent of disease stage. *Leukemia,* 26**,** 826-30.

BRITT-COMPTON, B., ROWSON, J., LOCKE, M., MACKENZIE, I., KIPLING, D. & BAIRD, D. M. (2006) Structural stability and chromosome-specific telomere length is governed by cis-acting determinants in humans. *Human molecular genetics,* 15**,** 725-33.

BROCCOLI, D. (2004) Function , replication and structure of the mammalian telomere. *Cytotechnology,* 45**,** 3-12.

BROCCOLI, D., SMOGORZEWSKA, A., CHONG, L. & DE LANGE, T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature Publishing Group,* 17**,** 231-235.

BROWN, S & BICKNELL R, (2001) Hypoxia and oxidative stress in breast cancer Oxidative stress: its effects on the growth, metastatic potential and response to therapy of breast *cancer Breast Cancer Research,* 3:323–327

BROWN W., MACKINNON P., VILLASANTE A., SPURR N., BUCKLE V., DOBSON M. (1990). Structure and polymorphism of human telomere-associated DNA. *Cell.;*63(1):119-132.

BRUIN, D. D., KANTROW, S. M., LIBERATORE, R. A. & ZAKIAN, V. A. (2000) Telomere Folding Is Required for the Stable Maintenance of Telomere Position Effects in Yeast Telomere Folding Is Required for the Stable Maintenance of Telomere Position Effects in Yeast. *Molecular and cellular biology,* 20**,** 7991.

BRYANT TM, ENGLEZOU A, DALLA-POZZA L, DUNHAM MA, REDDEL RR. (1997). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.* 3:1271–74

BRYANT TM, REDDEL RR. (1994). SV40-induced immortalization of human cells. *Crit. Rev. Oncol.* 5:331–57

BRYANT, H. E., SCHULTZ, N., THOMAS, H. D., PARKER, K. M., FLOWER, D., LOPEZ, E., KYLE, S., MEUTH, M., CURTIN, N. J. & HELLEDAY, T. (2005) Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature,* 434**,** 913-7.

CALDECOTT, K. W. (2008) Single-strand break repair and genetic disease. *Nature reviews. Genetics,* 9**,** 619-31.

CALDERON-MARGALIT, R. & PALTIEL, O. (2004) Prevention of breast cancer in women who carry BRCA1 or BRCA2 mutations: a critical review of the literature. *International journal of cancer. Journal international du cancer,* 112**,** 357-64.

CALLINAN, A. & BATZER, M. A. (2006) Retrotransposable Elements and Human Disease. *Genome and Disease,* 1**,** 104-115.

CAMPISI, J. The Biology of Replicative Senescence. *Journal of Biological Chemistry,* 49.

CAMPISI, J. & D'ADDA DI FAGAGNA, F. (2007) Cellular senescence: when bad things happen to good cells. *Nature reviews. Molecular cell biology,* 8**,** 729-40.

CAPPER, R., BRITT-COMPTON, B., TANKIMANOVA, M., ROWSON, J., LETSOLO, B., MAN, S., HAUGHTON, M. & BAIRD, D. M. (2007) The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes & development***,** 2495-2508.

CAPPER, R., BRITT-COMPTON, B., TANKIMANOVA, M., ROWSON, J., LETSOLO, B., MAN, S., HAUGHTON, M. & BAIRD, D. M. (2007) The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes & development,* 21**,** 2495-508.

CARDOSO, F., PICCART-GEBHART, M. & VAN, L. (2007) The MINDACT trial : The first prospective clinical validation of a genomic tool. *Molecular oncology,* 1**,** 246-251.

CASJENS, S. (1999) Evolution of the linear DNA replicons of the Borrelia spirochetes. *Current*

opinion in microbiology, 2**,** 529-34.

CAWTHON, R. M. (2002) Telomere measurement by quantitative PCR. *Nucleic acids research,* 30**,** 1-6.

CAWTHON, R. M., SMITH, K. R., O'BRIEN, E., SIVATCHENKO, A. & KERBER, R. A. (2003) Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet,* 361**,** 393-5.

CECH TR. 2004. Beginning to understand the end of the chromosome. Cell 116:273–79

CELLI, G. B. & DE LANGE, T. (2005) DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature cell biology,* 7**,** 712-8.

CELLI, G. B., DENCHI, E. L. & DE LANGE, T. (2006) Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. *Nature cell biology,* 8**,** 885-90.

CESARE, A. J., GRIFFITH, J. D., CESARE, A. J. & GRIFFITH, J. D. (2004) Telomeric DNA in ALT Cells Is Characterized by Free Telomeric Circles and Heterogeneous t-Loops Telomeric DNA in ALT Cells Is Characterized by Free Telomeric Circles and Heterogeneous t-Loops. *Molecular and cellular biology,* 24**,** 9948.

CESARE, A. J., QUINNEY, N., WILLCOX, S., SUBRAMANIAN, D. & GRIFFITH, J. D. (2003) Telomere looping in P. sativum (common garden pea). *The Plant Journal,* 36**,** 271-279.

CHAI, W., SFEIR, A. J., HOSHIYAMA, H., SHAY, J. W. & WRIGHT, W. E. (2006) The involvement of the Mre11/Rad50/Nbs1 complex in the generation of G-overhangs at human telomeres. *EMBO reports,* 7**,** 225-30.

CHAI, W., SHAY, J. W. & WRIGHT, W. E. (2005) Human Telomeres Maintain Their Overhang Length at Senescence †. *Molecular and cellular biology,* 25**,** 2158-2168.

CHADENEAU C, HAY K, HIRTE HW, GALLINGER S, BACCHETTI S. (1995)Telomerase activity associated with acquisition of malignancy in human colorectal cancer. *Cancer Res*;55(12):2533-2536.

CHANG, S., MULTANI, A. S., CABRERA, N. G., NAYLOR, M. L., LAUD, P., LOMBARD, D., PATHAK, S., GUARENTE, L. & DEPINHO, R. A. (2004) Essential role of limiting telomeres in the pathogenesis of Werner syndrome. *Nature genetics,* 36**,** 877-82.

CHEAH, P., LOOI, L. (2001) p53: an overview of over two decades of study. *Malaysian Journal of Pathology.* Jun;23(1):9-16.

CHEN, C. W. (1996) Complications and implications of linear bacterial chromosomes. *Trends in*

genetics : TIG, 12**,** 192-6.

CHEN, X., YU, X., CHEN, J., ZHANG, Z., TUAN, J. & SHAO, Z. (2013) Analysis in Early Stage Triple-Negative Breast Cancer Treated With Mastectomy Without Adjuvant Radiotherapy : Patterns of Failure and Prognostic Factors. *Cancer***,** 1-9.

CHEUNG, I., SCHERTZER, M., BAROSS, A., ROSE, A. M., LANSDORP, P. M. & BAIRD, D. M. (2004) Strain-specific telomere length revealed by single telomere length analysis in Caenorhabditis elegans. *Nucleic acids research,* 32**,** 3383-91.

CHIN, K., SOLORZANO, C. O. D., KNOWLES, D., JONES, A., CHOU, W., RODRIGUEZ, E. G., KUO, W.-L., LJUNG, B.-M., CHEW, K., MYAMBO, K., MIRANDA, M., KRIG, S., GARBE, J., STAMPFER, M., YASWEN, P., GRAY, J. W. & LOCKETT, S. J. (2004) In situ analyses of genome instability in breast cancer. *Nature genetics,* 36**,** 984-988.

CHIN, L., ARTANDI, S. E., SHEN, Q., TAM, A., LEE, S. L., GOTTLIEB, G. J., GREIDER, C. W. & DEPINHO, R. A. (1999) P53 Deficiency Rescues the Adverse Effects of Telomere Loss and Cooperates With Telomere Dysfunction To Accelerate Carcinogenesis. *Cell,* 97**,** 527-38.

CHONG, L. V. S., BAS; BROCCOLI, DOMINIQUE; ERDJUMENT-BROMAGE, HEDIYE; HANISH, JOHN; TEMPST, PAUL AND DE LANGE, TITIA (1991) A human telomeric protein. *Science,* 270**,** 1663-1667.

CLARKE, C. L. (2006).Mapping loss of heterozygosity in normal human breast cells from BRCA1/2 carriers. *Br.J. Cancer* 95, 515–519

COHN, M. & BLACKBURN, E. H. (1994) Telomerase in Yeast. *Science,* 269**,** 396-400.

COLEMAN, J., BAIRD, D. M. & ROYLE, N. J. (1999) The plasticity of human telomeres demonstrated by a hypervariable telomere repeat array that is located on some copies of 16p and 16q. *Human molecular genetics,* 8**,** 1637-46.

COLLADO, M. & SERRANO, M. (2010) Senescence in tumours: evidence from mice and humans. *Nature reviews. Cancer,* 10**,** 51-7.

CONG, Y.-S., WRIGHT, W. E. & SHAY, J. W. (2002) Human Telomerase and Its Regulation. *Microbiology and Moilecular Biology reviews,* 66**,** 407-425.

CONRAD, M. N., WRIGHT, J. H., WOLF, A. J. & ZAKIAN, V. A. (1990) RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability. *Cell,* 63**,** 739-50.

COOK, B. D., DYNEK, J. N., CHANG, W., SHOSTAK, G. & SMITH, S. (2002) Role for the Related Poly (ADP-Ribose) Polymerases Tankyrase 1 and 2 at Human Telomeres. *Molecular and cellular biology,* 22**,** 332-342.

COOPER, J.P., NIMMO, E.R., ALLSHIRE, R.C., AND CECH, T.R. (1997). Regulation of

telomere length and function by a Myb-domain protein in fission yeast. *Nature* 385, 744–747.

CORTES-LEDESMA, F., A (2006) Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange. *EMBO reports,* 7.

CORTEZ, D. (1999) Requirement of ATM-Dependent Phosphorylation of Brca1 in the DNA Damage Response to Double-Strand Breaks. *Science,* 286**,** 1162-1166.

COUNTER, C. M., AVILION, A. A., LEFEUVREL, C. E., STEWART, N. G., GREIDER, C. W., HARLEY, C. B. & BACCHETTIL, S. (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *The EMBO journal,* 1**,** 1921-1992.

COURT, R., CHAPMAN, L., FAIRALL, L. & RHODES, D. (2005) How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO reports,* 6**,** 39-45.

CURTIS RE, BOICE JD JR, STOVALL M (1989). Leukemia risk following radiotherapy for breast cancer. *J Clin Oncol 7*(1):21-29.

CURTIS RE, BOICE JD JR, STOVALL M (1992). Risk of leukemia after chemotherapy and radiation treatment for breast cancer. *N Engl J Med* 326(26):1745-1751.

CURTIS RE, BOICE JD JR, SHRINER DA (1996). Second cancers after adjuvant tamoxifen therapy for breast cancer. *J Natl Cancer Inst* 88(12):832-834.

CURTIS RE, FREEDMAN DM, SHERMAN ME, (2004). Risk of malignant mixed mullerian tumors after tamoxifen therapy for breast cancer. *J Natl Cancer Inst* 96(1):70-74.

CURTIS, C; SHAH, S, CHIN, S-F, TURASHVILI, G (2012). The genomic and transcriptomic architechture of 2,000 breast tumours reveals novel subgroups. *Nature*. 486:7403; 346-352

DAVALOS, A. R., COPPE, J.-P., CAMPISI, J. & DESPREZ, P.-Y. (2010) Senescent cells as a source of inflammatory factors for tumor progression. *Cancer metastasis reviews,* 29**,** 273-83.

DAVIDOFF, AM., (1991). Genetic basis for p53 overexpression in human breast cancer. *Proc*

Natl Acad Sci USA, 88, pp.5006–5010.

DAVIS T, BAIRD DM, HAUGHTON MF, JONES CJ, KIPLING D, (2005) Prevention of accelerated cell aging in Werner syndrome using a p38 mitogen-activated protein kinase inhibitor.*J Gerontol A Biol Sci Med Sci*, Volume 60, 11 pp.1386-1393

DE BRUIN, D., ZAMAN, Z., LIBERATORE, R. A. & PTASHNE, M. (2001) Telomere looping permits gene activation by a downstream UAS in yeast. *Nature,* 409**,** 109-13.

DE LANGE, T. (1995). Telomere dynamics and genome instability in human cancer. In: Telomeres, ed. E.H.B.a.C.W. Greider, *Cold Spring Harbor, NY: Cold Spring Harbor* Press, 265–293.

DE LANGE, T. (2002). Protection of mammalian telomeres. *Oncogene* 21, 532–540.

DE LANGE, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development,* 19**,** 2100-10.

DE LANGE, T. (2005) Telomere-related genome instability in cancer. *Cold Spring Harbor symposia on quantitative biology,* 70**,** 197-204.

DE LANGE, T. (2010) Telomere biology and DNA repair: enemies with benefits. *FEBS letters,* 584**,** 3673-4.

DE LANGE, T., SHIUE, L., MYERS, R. M., COX, D. R., NAYLOR, S. L., KILLERY, A. M. & VARMUS, H. E. (1990) Structure and variability of human chromosome ends. *Molecular and cellular biology,* 10**,** 518-27.

DE LEON, J., SUSCE, M., JOHNSON, M., HARDIN, M., MAW, L., SHAO, A., ALLEN, A., CHIAFARI, F., HILLMAN, G., NIKOLOFF, D. (2009).DNA microarray technology in the clinical environment: the AmpliChip CYP450 test for CYP2D6 and CYP2C19 genotyping. *CNS Spectrum*. Jan;14(1):19-34

DE VOS, P., SCHMITT, J., VERHOEVEN, G. & STUNNENBERG, H. G. (1994) Human androgen receptor expressed in HeLa cells activates transcription in vitro. *Nucleic acids research,* 22**,** 1161-6.

DEBACKER, K. & KOOY, R. F. (2007) Fragile sites and human disease. *Human molecular genetics,* 16 Spec No**,** R150-8.

DEININGER, P. L. & BATZER, M. A. (1999) Alu repeats and human disease. *Molecular*

genetics and metabolism, 67**,** 183-93.

DENG, Y., GUO, X., FERGUSON, D. O. & CHANG, S. (2009) Multiple roles for MRE11 at uncapped telomeres. *Nature,* 460**,** 914-8.

DENG, Z., NORSEEN, J., WIEDMER, A., RIETHMAN, H. & LIEBERMAN, P. M. (2009) TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Molecular Cell,* 35**,** 403-413.

DENT, R., TRUDEAU, M., PRITCHARD, K. I., HANNA, W. M., KAHN, H. K., SAWKA, C. A., LICKLEY, L. A., RAWLINSON, E., SUN, P. & NAROD, S. A. (2007) Triple-Negative Breast Cancer : Clinical Features and Patterns of Recurrence riple-Negative Breast Cancer : Clinical Features and Patterns of Recurrence. *Clinical Cancer Research,* 13**,** 4429-4434.

DIMRI, G. P., LEE, X., BASILE, G., ACOSTA, M., SCOTT, G., ROSKELLEY, C., MEDRANO, E. E., LINSKENS, M., RUBELJ, I. & PEREIRA-SMITH, O. (1995) A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 9363-7.

DING, Z., WU, C.-J., JASKELIOFF, M., IVANOVA, E., KOST-ALIMOVA, M., PROTOPOPOV, A., CHU, G. C., WANG, G., LU, X., LABROT, E. S., HU, J., WANG, W., XIAO, Y., ZHANG, H., ZHANG, J., ZHANG, J., GAN, B., PERRY, S. R., JIANG, S., LI, L., HORNER, J. W., WANG, Y. A., CHIN, L. & DEPINHO, R. A. (2012) Telomerase reactivation following telomere dysfunction yields murine prostate tumors with bone metastases. *Cell,* 148**,** 896-907.

DOBZHANSKY, T (1946) Genetics of natural populations: Xiii. Recombination and variability in populations of Drosophila Pseudoobscura. *Genetics* 31:269-290,

DUNHAM, M. A., NEUMANN, A. A., FASCHING, C. L. & REDDEL, R. R. (2000) Telomere maintenance by recombination in human cells. *Nature genetics,* 26**,** 447-50.

DUNNWALD, L. K., ROSSING, M. A. & LI, C. I. (2007) Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast cancer research : BCR,* 9**,** R6.

EDWARDS, R. A., OLSEN, G. J. & MALOY, S. R. (2002) Comparative genomics of closely related salmonellae. *Trends in microbiology,* 10**,** 94-9.

EPSTEIN, C., MARTIN, G., SCHULTZ, A., MOTULSKY, A. (1966)Werner's syndrome a review of its symptomatology, natural history, pathologic features, genetics and relationship to the natural aging process. *Medicine (Baltimore*). May;45(3):177-221

ESPEJEL, S., MARTÍN, M., KLATT, P., MARTÍN-CABALLERO, J., FLORES, J. M. & BLASCO, M. A. (2004) Shorter telomeres, accelerated ageing and increased lymphoma in DNA-PKcs-deficient mice. *EMBO reports,* 5**,** 503-9.

ESPINA, V. & LIOTTA, L. A. (2011) What is the malignant nature of human ductal carcinoma in situ? *Nature reviews. Cancer,* 11**,** 68-75.

FARR, C., FANTES, J., GOODFELLOW, P., AND COOKE, H. (1991). Functional reintroduction of human telomeres into mammalian cells. *Proc. Natl. Acad. Sci. USA* 88, 7006–7010.

FERLINI C, SCAMBIA G, MARONE M, DISTEFANO M, GAGGINI C, FERRANDINA G, FATTOROSSI A, ISOLA G, BENEDETTI PANICI P, MANCUSO S: (1999) Tamoxifen induces oxidative stress and apoptosis in oestrogen receptor-negative human cancer cell lines. *Br J Cancer*, 79:257-263

FERRÓN, S., MIRA, H., FRANCO, S., CANO-JAIMEZ, M., BELLMUNT, E., RAMÍREZ, C., FARIÑAS, I. & BLASCO, M. A. (2004) Telomere shortening and chromosomal instability abrogates proliferation of adult but not embryonic neural stem cells. *Development (Cambridge, England),* 131**,** 4059-70.

FILIPPO, J. S., SUNG, P. & KLEIN, H. (2008) Mechanism of Eukaryotic Homologous Recombination. *Annual review of cell and developmental biology,* 77**,** 229-260.

FISHEL, R., EWEL, A. & LESCOE, M. K. (1994) Purified Human MSH2 Protein Binds to DNA Containing Mismatched Nucleotides Purified Human MSH2 Protein Binds to DNA Containing MismatchedNucleotid &. *Cancer research,* 54**,** 5539-5542.

FLINT, J., WILKIE, A., BUCKLE, V., WINTER, R., HOLLAND, A. & MCDERMID, H. (1995) A detection of subtelomeric chromosomal rearrangements in idiopathic mental rearrangements. *Nature Publishing Group,* 9**,** 132-141.

FOOTE, F. & STEWART, F. (1941) Lobular carcinoma in situ: a rare form of mammary cancer. *American journal of pathology,* 17**,** 491-496.

FOOTE FW Jr, STEWART FW. (1946) .A histologic classification of carcinoma of the breast. *Surgery.* Jan;19:74-99

FRIEDBURG, E., WALKER, G., SIEDE, W., SCHULTZ, R. (2006). DNA repair and mutagenesis (2nd ed.). *ASM Press*. ISBN 978-1-55581-319-2.

FUTREAL, P. A., SÖDERKVIST, P., MARKS, J. R., SÃ, P., IGLEHART, J. D., COCHRAN, C., BARRETT, J. C. & WISEMAN, R. W. (1992) Detection of Frequent Allelic Loss on Proximal Chromosome 17q in Sporadic Breast Carcinoma Using Microsatellite Length Polymorphisms. *Cancer research,* 52**,** 2624-2627.

GARVIK, B., CARSON, M. & HARTWELL, L. (1995) Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint. *Molecular and cellular biology,* 15**,** 6128-38.

GASCO, M., SHAMI, S. & CROOK, T. (2002) The p53 pathway in breast cancer. *Breast Cancer Research*.

GELMINI, S., POGGESI, M., DISTANTE, V., BIANCHI, S., SIMI, L., LUCONI, M., CASINI, C., CATALIOTTI, L., PAZZAGLI, M. & ORLANDO, C. (2004) Tankyrase , a positive regulator of telomere elongation , is over expressed in human breast cancer. *Cancer Letters,* 216**,** 81-87.

GERTLER, R., ROSENBERG, R., STRICKER, D., FRIEDERICHS, J., HOOS, A., WERNER, M., ULM, K., HOLZMANN, B., NEKARDA, H. & SIEWERT, J.-R. (2004) Telomere length and human telomerase reverse transcriptase expression as markers for progression and prognosis of colorectal carcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* 22**,** 1807-14.

GISSELSSON D, HÖGLUND M, MERTENS F, JOHNSSON B, DAL CIN P, VAN DEN BERGHE H (1999). The structure and dynamics of ring chromosomes in human neoplastic and non-neoplastic cells. *Hum Genet*.;104:315–325.

GISSELSSON, D., JONSON, T., PETERSÉN, A., STRÖMBECK, B., DAL CIN, P., HÖGLUND, M., MITELMAN, F., MERTENS, F. & MANDAHL, N. (2001) Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 12683-8.

GORDON, K. E., IRELAND, H., ROBERTS, M., STEEGHS, K., MCCAUL, J. A., MACDONALD, D. G. & PARKINSON, E. K. (2003) High levels of telomere dysfunction bestow a selective disadvantage during the progression of human oral squamous cell carcinoma. *Cancer research,* 63**,** 458-67.

GOTTLICH B., REICHENBERGER,S., FELDMANN,E. AND PFEIFFER,P. (1998) Rejoining of DNA double-strand breaks in vitro by single-strand annealing. *Eur. J. Biochem.*, 258, 387–395.

GOYTISOLO, F. A. & BLASCO, M. A. (2002) Many ways to telomere dysfunction : in vivo studies using mouse models. *Nature Publishing Group***,** 584-591.

GRAHAM, K., GE, X. & MORENAS, A. D. (2011) Gene Expression Profiles of Estrogen Receptor − Positive and Estrogen Receptor − Negative Breast Cancers Are Detectable in Histologically Normal Breast Epithelium. *Clinical Cancer Research,* 17**,** 236-246.

GRANDIN, N. & CHARBONNEAU, M. (2008) Protection against chromosome degradation at the telomeres. *Biochimie,* 90**,** 41-59.

GREENBURG, R. A. (2005) Telomeres, crisis and cancer. *Current molecular medicine,* 5**,** 213-8.

GREIDER, C. W. (1996) Telomere length regulation. *Annual review of biochemistry,* 65**,** 337-65.

GREIDER, C. W. (1999) Telomeres Do D-Loop – T-Loop Minireview. *Cell,* 97**,** 419-422.

GRIFFITH, J., BIANCHI, A. & DE LANGE, T. (1998) TRF1 promotes parallel pairing of telomeric tracts in vitro. *Journal of molecular biology,* 278**,** 79-88.

GRIFFITH, J. K., BRYANT, J. E., FORDYCE, C. A., GILLILAND, F. D., JOSTE, N. E. & MOYZIS, R. K. (1999) Reduced telomere DNA content is correlated with genomic instability and metastasis in invasive human breast carcinoma. *Breast cancer research and treatment,* 54**,** 59-64.

GUO, A., SALOMONI, P., LUO, J., SHIH, A., ZHONG, S., GU, W. & PANDOLFI, P. P. (2000) The function of PML in p53-dependent apoptosis. *Nature cell biology,* 2**,** 730-6.

HAIGIS, K. M., HOFF, P. D., WHITE, A., SHOEMAKER, A. R., HALBERG, R. B. & DOVE, W. F. (2004) Tumor regionality in the mouse intestine reflects the mechanism of loss of Apc function. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 9769-73.

HAMPTON, O. A., DEN HOLLANDER, P., MILLER, C. A., DELGADO, D. A., LI, J., COARFA, C., HARRIS, R. A., RICHARDS, S., SCHERER, S. E., MUZNY, D. M., GIBBS, R. A., LEE, A. V. & MILOSAVLJEVIC, A. (2009) A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. *Genome research,* 19**,** 167-77.

HARLEY, C., FUTCHER, B. & GREIDER, C. W. (1990) Telomeres shorten during ageing of human fibroblasts. *Nature,* 345**,** 458.

HARLEY, C., VAZIRI, H., COUNTER, C., ALLSOPP, R., (1992)The telomere hypothesis of cellular aging. *Experimental Gerontology* Jul-Aug;27(4):375-82.

HÄSLER, J. & STRUB, K. (2006) Alu elements as regulators of gene expression. *Nucleic acids research,* 34**,** 5491-7.

HAYFLICK L, MOORHEAD, PS (1961). "The serial cultivation of human diploid cell strains". *Exp Cell Res* 25 (3): 585–621.

HAYFLICK L. (1965). "The limited in vitro lifetime of human diploid cell strains". *Exp. Cell Res*. 37 (3): 614–636.

HEAPHY, C. M., BAUMGARTNER, K. B., BISOFFI, M., BAUMGARTNER, R. N. & GRIFFITH, J. K. (2007) Telomere DNA Content Predicts Breast Cancer − Free Survival Interval Survival Interval. *Clinical Cancer Research,* 13**,** 7037-7043.

HEAPHY, C. M., BAUMGARTNER, K. B. & GRIFFITH, J. K. (2010) Genomic instability

demonstrates similarity between DCIS and invasive carcinomas. *Breast cancer research and treatment,* 117**,** 17-24.

HEAPHY, C. M., SUBHAWONG, A. P., GROSS, A. L., KONISHI, Y., KOUPRINA, N., ARGANI, P., VISVANATHAN, K. & MEEKER, A. K. (2011) Shorter telomeres in luminal B, HER-2 and triple-negative breast cancer subtypes. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc,* 24**,** 194-200.

HELENA MANGS, A. & MORRIS, B. J. (2007) The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Current genomics,* 8**,** 129-36.

HELLER, T., KIRCHHEINER, J., ARMSTRONG, V., LUTHE, H., TZVETKOV, M., BROCKMOLLER, J., OELLERICH, M., (2006) AmpliChip CYP450 GeneChip: a new gene chip that allows rapid and accurate CYP2D6 genotyping. *Therapeutic Drug Monitoring* Oct;28(5):673-7.

HEMANN, M. T. & GREIDER, C. W. (2000) Wild-derived inbred mouse strains have short telomeres. *Nucleic acids research,* 28**,** 4474-8.

HEMANN, M. T., STRONG, M. A., HAO, L. Y. & GREIDER, C. W. (2001) The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell,* 107**,** 67-77.

HERNANDEZ-BOUSSARD, T., MONTESANO, R., HAINAUT, P. (1999) Analysis of somatic mutations of the p53 gene in human cancers: a tool to generate hypotheses about the natural history of cancer. *IARC Scientific* ;(146):43-53

HILSENBECK, S. R., P; MOOR, C; CHAMNESS, G; OSBORNE, K AND CLARK, G (1998) Time-dependence of Hazard ratios for prognostic factors inb primary breast cancer. *Breast cancer research and treatment,* 52**,** 227-237.

HINESTROSA, M. C., DICKERSIN, K., KLEIN, P., MAYER, M., NOSS, K., SLAMON, D., SLEDGE, G. & VISCO, F. M. (2007) Shaping the future of biomarker research in breast cancer to ensure clinical relevance. *Nature reviews. Cancer,* 7**,** 309-15.

HOEIJMAKERS, J. H. J. (2001) DNA repair mechanisms. *Maturitas,* 38**,** 17-22.

HOPP, T. A. (2004) Breast Cancer Patients with Progesterone Receptor PR-A-Rich Tumors Have Poorer Disease-Free Survival Rates. *Clinical Cancer Research,* 10**,** 2751-2760.

HOSOKAWA, Y. & ARNOLD, A. (1998) Mechanism of cyclin D1 (CCD1, PRAD1) Overexpression in Human Cancer Cells : Analysis of Allele-Specific Expression. *Genes, chromosomes & cancer,* 71**,** 66-71.

HUANG, T.-W. & CHEN, C. W. (2008) DNA polymerase I is not required for replication of linear chromosomes in streptomyces. *Journal of bacteriology,* 190**,** 755-8.

HUEBNER K., CROCE C.M. (2001) FRA3B and other common fragile sites: the weakest links. *Nat. Rev. Cancer.*;1:214–221.

HUGH, J., HANSON, J., CHEANG, M. C. U., NIELSEN, T. O., PEROU, C. M., DUMONTET, C., REED, J., KRAJEWSKA, M., TREILLEUX, I., RUPIN, M., MAGHERINI, E., MACKEY, J., MARTIN, M. & VOGEL, C. (2009) Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use of an immunohistochemical definition in the BCIRG 001 trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* 27**,** 1168-76.

ILIAN, C. & HUIRE, M. (2006) Telomere and Telomerase : brief review of a history initiated by Hermann Müller and Barbara McClintock. *Colombia Medica,* 37**,** 336-339.

INDIVIGLIO, S. M. & BERTUCH, A. A. (2009) Ku's essential role in keeping telomeres intact. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 12217-8.

JACKSON, S. P. & BARTEK, J. (2009) The DNA-damage response in human biology and disease. *Nature,* 461**,** 1071-8.

JENSEN, R. B., CARREIRA, A. & KOWALCZYKOWSKI, S. C. (2010) Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature,* 467**,** 678-83.

JONES, C. H., PEPPER, C. & BAIRD, D. M. (2012) Telomere dysfunction and its role in haematological cancer. *British journal of haematology,* 156**,** 573-87.

JUMAS-BILAK, E., MICHAUX-CHARACHON, S., BOURG, G., RAMUZ, M. & ALLARDET-SERVENT, A. (1998) Unconventional genomic organization in the alpha subgroup of the Proteobacteria. *Journal of bacteriology,* 180**,** 2749-55.

KAELIN, W (2005) The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer* 5:689-698

KARLSEDER, J. (1999) p53- and ATM-Dependent Apoptosis Induced by Telomeres Lacking TRF2. *Science,* 283**,** 1321-1325.

KARLSEDER, J. (1999) p53- and ATM-Dependent Apoptosis Induced by Telomeres Lacking TRF2. *Science,* 283**,** 1321-1325.

KARLSEDER, J., HOKE, K., MIRZOEVA, O. K., BAKKENIST, C., KASTAN, M. B. & PETRINI, J. H. J. (2004) The Telomeric Protein TRF2 Binds the ATM Kinase and Can Inhibit the ATM-Dependent DNA Damage Response. *PLoS biology,* 2.

KELCHNER SA, WENDEL JF (1996) Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Curr Genet* 30: 259–262.

KIM, S. H., KAMINKER, P. & CAMPISI, J. (1999) TIN2, a new regulator of telomere length in human cells. *Nature genetics,* 23**,** 405-12.

KIM, S.-H., BEAUSEJOUR, C., DAVALOS, A. R., KAMINKER, P., HEO, S.-J. & CAMPISI, J. (2004) TIN2 mediates functions of TRF2 at human telomeres. *The Journal of biological chemistry,* 279**,** 43799-804.

KIPLING, D and COOKE, H. (1990) Hypervariable ultra-long telomeres in mice. *Nature,* 347**,** 400-402.

KLEMI, P. J., TOIKKANEN, S., PARVINEN, I. & JOENSUU, H. (1997) Mammography screening interval and the frequency of interval cancers in a population-based screening. *British journal of cancer,* 75**,** 762-766.

KLOBUTCHER, L. A. & RUDDLE, F. H. (1981) Chromosome mediated gene transfer. *Annual review of biochemistry,* 50**,** 533-54.

KLOBUTCHER, L. A., SWANTON, M. T., DONINI, P. & PRESCOTT, D. M. (1981) All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proceedings of the National Academy of Sciences of the United States of America,* 78**,** 3015-9.

KOLQUIST, K. A., ELLISEN, L. W., COUNTER, C. M., MEYERSON, M., TAN, L. K., WEINBERG, R. A., HABER, D. A. & GERALD, W. L. (1998) Expression of TERT in early premalignant lesions and a subset of cells in normal tissues. *Nature genetics,* 19**,** 182-6.

KORNBERG, A.(1974) DNA Synthesis *W.H. Freeman,* San Francisco

KRAUS, W. L. & LIS, J. T. (2003) PARP Goes Transcription. *Cell,* 113**,** 677-683.

KRAUSKOPF, A. A. B., E (1996) Control of telomere growth by interactions of RAP1 with the most distal telomeric repeats. *Nature,* 386**,** 354-357.

KULKARNI, A., ZSCHENKER, O., REYNOLDS, G., MILLER, D. & MURNANE, J. P. (2010) Effect of telomere proximity on telomere position effect, chromosome healing, and sensitivity to DNA double-strand breaks in a human tumor cell line. *Molecular and cellular biology,* 30**,** 578-89.

KWEI, K. A., KUNG, Y., SALARI, K., HOLCOMB, I. N. & POLLACK, J. R. (2010) Genomic instability in breast cancer: pathogenesis and clinical implications. *Molecular oncology,* 4**,** 255-66.

KWEI, K. K., YVONNE, HOLCOMB, ILOINA AND POLLOCK, JONATHON (2011) Genomic instability in breast cancer: pathogenesis and clinical implications. *Molecular oncology,* 4**,** 255-266.

KYRION, G., LIU, K. & LUSTIG, A. J. (1993) RAP1 and telomere structure regulate telomere position effects in Saccharomyces cerevisiae. *Genes & development,* 7**,** 1146-1159.

LANGE, T. D. (2005) Shelterin : the protein complex that shapes and safeguards human telomeres. *Genes & development,* 19**,** 2100-2110.

LANSDORP, P. M., VERWOERD, N. P., VAN DE RIJKE, F. M., DRAGOWSKA, V., LITTLE, M. T., DIRKS, R. W., RAAP, A. K. & TANKE, H. J. (1996) Heterogeneity in telomere length of human chromosomes. *Human molecular genetics,* 5**,** 685-91.

LAURENSON, P. & RINE, J. (1992) Silencers , Silencing , and Heritable Transcriptional States. *Microbiological reviews,* 56**,** 543-560.

LEE, H., TRAINER, A. H., FRIEDMAN, L. S., THISTLETHWAITE, F. C., EVANS, M. J., PONDER, B. A. J. & VENKITARAMAN, A. R. (1999) Transformation in Cells Lacking the Breast Cancer Susceptibility Gene , Brca2. *Molecular Cell,* 4**,** 1-10.

LEE, H.-H., HSU, C.-C., LIN, Y.-L. & CHEN, C. W. (2011) Linear plasmids mobilize linear but not circular chromosomes in Streptomyces: support for the 'end first' model of conjugal transfer. *Microbiology (Reading, England),* 157**,** 2556-68.

LEE, H. W., BLASCO, M. A., GOTTLIEB, G. J., HORNER, J. W., GREIDER, C. W. & DEPINHO, R. A. (1998) Essential role of mouse telomerase in highly proliferative organs. *Nature,* 392**,** 569-74.

LEE, J.-H. & PAULL, T. T. (2007) Activation and regulation of ATM kinase activity in response to DNA double-strand breaks. *Oncogene,* 26**,** 7741-8.

LEE, K. H., RUDOLPH, K. L., JU, Y. J., GREENBERG, R. A., CANNIZZARO, L., CHIN, L., WEILER, S. R. & DEPINHO, R. A. (2001) Telomere dysfunction alters the chemotherapeutic profile of transformed cells. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 3381-6.

LEE, Y., MANGASARIAN, O. L. & WOLBERG, W. H. Breast Cancer Survival Analysis and Chemotherapy via Generalized Support Vector Machines.

LEES-MILLER, S. P. & MEEK, K. (2003) Repair of DNA double strand breaks by non-homologous end joining. *Biochimie,* 85**,** 1161-1173.

LEE-THEILEN, M., MATTHEWS, A. J., KELLY, D., ZHENG, S. & CHAUDHURI, J. (2011) CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nature structural & molecular biology,* 18**,** 75-9.

LEKOMTSEV, S., GUIZETTI, J., POZNIAKOVSKY, A., GERLICH, D. W. & PETRONCZKI,

M. (2010) Evidence that the tumor-suppressor protein BRCA2 does not regulate cytokinesis in human cells. *Journal of cell science,* 123**,** 1395-400.

LEONI, C., VOLPICELLA, M., DE LEO, F., GALLERANI, R. & CECI, L. R. (2011) Genome walking in eukaryotes. *The FEBS journal,* 278**,** 3953-77.

LETSOLO, B. T., ROWSON, J. & BAIRD, D. M. (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**,** 1841-52.

LI, B. & LANGE, T. D. (2003) Rap1 Affects the Length and Heterogeneity of Human Telomeres. *Molecular biology of the Cell,* 14**,** 5060-5068.

LI, X. & HEYER, W.-D. (2008) Homologous recombination in DNA repair and DNA damage tolerance. *Cell research,* 18**,** 99-113.

LI, Y., NIE, H., WU, D., ZHANG, J., WEI, X. & YING, W. (2013) Poly(ADP-ribose) polymerase mediates both cell death and ATP decreases in SIRT2 inhibitor AGK2-treated microglial BV2 cells. *Neuroscience letters***,** 1-5.

LIN, T. T., LETSOLO, B. T., JONES, R. E., ROWSON, J., PRATT, G., FEGAN, C., PEPPER, C. & BAIRD, D. M. (2010) Telomere dysfunction and fusion during the progression of a human malignancy. *Blood,* 44.

LIN, T. T., LETSOLO, B. T., JONES, R. E., ROWSON, J., PRATT, G., HEWAMANA, S., FEGAN, C., PEPPER, C. & BAIRD, D. M. (2010) Telomere dysfunction and fusion during the progression of chronic lymphocytic leukemia: evidence for a telomere crisis. *Blood,* 116**,** 1899-907.

LIN, T. T., LETSOLO, B. T., JONES, R. E., ROWSON, J., PRATT, G., HEWAMANA, S., FEGAN, C., PEPPER, C. & BAIRD, D. M. (2010) Telomere dysfunction and fusion during the progression of chronic lymphocytic leukemia: evidence for a telomere crisis. *Blood,* 116**,** 1899-907.

LIU, D., CONNOR, M. S. O., QIN, J. & SONGYANG, Z. (2004) Telosome , a Mammalian Telomere-associated Complex Formed by Multiple Telomeric Proteins *. *The Journal of biological chemistry,* 279**,** 51338-51342.

LIU, Y., SNOW, B. E., HANDE, M. P., YEUNG, D., ERDMANN, N. J., WAKEHAM, A., ITIE, A., SIDEROVSKI, D. P., LANSDORP, P. M., ROBINSON, M. O. & HARRINGTON, L. (2000) The telomerase reverse transcriptase is limiting and necessary for telomerase function in vivo. *Current biology : CB,* 10**,** 1459-62.

LO AW, SABATIER L, FOULADI B, POTTIER G, RICOUL M, MURNANE JP. (2002) DNA amplification by breakage/fusion/ bridge cycles initiated by spontaneous telomere loss in a human cancer cell line. *Neoplasia.* 2002;4(6):531-538.

LOCKE, I. (2006) Loss of heterozygosity at the BRCA1 and BRCA2 loci detected in ductal lavage fluid from BRCA gene mutation carriers and controls. *Cancer Epidemiol. Biomarkers Prev*. 15, 1399–1402

LODISH H, BERK A, MATSUDAIRA P, KAISER CA, KRIEGER M, SCOTT MP, ZIPURSKY SL, DARNELL J. (2004). *Molecular Biology of the Cell*, p963. WH Freeman: New York, NY. 5th ed.

LOVEJOY, C, LI, W, REISENWEBER, S, DELANGE T (2012) Loss of ATRX, Genome Instability, and an Altered DNA Damage Response Are Hallmarks of the Alternative Lengthening of Telomeres Pathway *PLoS Genet* 8(7)

LU, J., STEEG, P. S., PRICE, J. E., KRISHNAMURTHY, S., MANI, S. A., REUBEN, J., CRISTOFANILLI, M., DONTU, G., BIDAUT, L., VALERO, V., HORTOBAGYI, G. N. & YU, D. (2009) Breast Cancer Metastasis : Challenges and Opportunities Breast Cancer Metastasis : Challenges and Opportunities. 4951-4953.

LU, L., ZHANG, C., ZHU, G., IRWIN, M., RISCH, H., MENATO, G., MITIDIERI, M., KATSAROS, D. & YU, H. (2011) Telomerase expression and telomere length in breast cancer and their associations with adjuvant treatment and disease outcome. *Breast cancer research : BCR,* 13**,** R56.

LUKE, B. & LINGNER, J. (2009) TERRA: telomeric repeat-containing RNA. *The EMBO journal,* 28**,** 2503-10.

LUSTIG, A. J., KURTZ, S. & SHORE, D. (1990) Involvement of the Silencer and UAS Binding Protein RAPI in Regulation of Telomere Length Pthly tchiof. *Science*, 250**,** 549-553.

MAELE, B. C., KARIN; TRETLI, STEINAR; AKSLEN, LARS AND GROTMOL, TOM (2009) Estrogen receptor B - an independant prognostic marker in estrogen receptor a and progesterone receptor-positive breast cancer? *APMIS,* 117**,** 644-650.

MACIEJEWSKI JP, TIU RV, O'KEEFE C (2009). Application of array-based whole genome scanning technologies as a cytogenetic tool in haematological malignancies. *Br J Haematol.;146*:479–488.

MAKAROV, V. L., HIROSE, Y. & LANGMORE, J. P. L. G. (1997) Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell,* 88**,** 657-666.

MALLEY, R. C. O., ALONSO, J. M., KIM, C. J., LEISSE, T. J. & ECKER, J. R. (2010) An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the Arabidopsis genome. *Nature Publishing Group*.

MARNETT, L. J. & JOHN, P. (2001) Endogenous DNA damage and mutation. *Trends in genetics : TIG,* 17**,** 214-221.

MARTENS, U. M. Z., MARK; POON, STEVEN; DRAGOWSKA, WIESLAWA; YUI, JANE; CHAVEZ, ELISABETH; WARD, RABAB AND LANSDORP, PETER (1998) Short telomeres on human chromosome 17p. *Nature genetics,* 18**,** 76-80.

MARTINEZ, P., THANASOULA, M., CARLOS, A. R., GÓMEZ-LÓPEZ, G., TEJERA, A. M., SCHOEFTNER, S., DOMINGUEZ, O., PISANO, D. G., TARSOUNAS, M. & BLASCO, M. A. (2010) Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites. *Nature cell biology,* 12**,** 768-80.

MARTINEZ-DELGADO, B., YANOWSKY, K., INGLADA-PEREZ, L., DOMINGO, S., URIOSTE, M., OSORIO, A. & BENITEZ, J. (2011) Genetic anticipation is associated with telomere shortening in hereditary breast cancer. *PLoS genetics,* 7**,** e1002182.

MARTINEZ-DELGADO, B., YANOWSKY, K., INGLADA-PEREZ, L., HOYA, M. D., CALDES, T., VEGA, A., BLANCO, A., MARTIN, T., GONZALEZ-SARMIENTO, R., BLASCO, M., ROBLEDO, M., URIOSTE, M., SONG, H., PHAROAH, P. & BENITEZ, J. (2012) Shorter telomere length is associated with increased ovarian cancer risk in both familial and sporadic cases. *Cancer genetics,* 49**,** 341-344.

MASON, D. J. & POWELSON, D. M. (1955) Nuclear division as observed in live bacteria by a new technique. 474-479.

MASSON, J. Y., TARSOUNAS, M. C., STASIAK, A. Z., STASIAK, A., SHAH, R., MCILWRAITH, M. J., BENSON, F. E. & WEST, S. C. (2001) Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes & development,* 15**,** 3296-307.

MCCLINTOCK, B. (1931).Cytological observations of deficiencies involving known genes, translocations and an inversion in Zea mays. *Missouri Agr. Exp. Sta. Res. Bull.* 163,1–48

MCCLINTOCK, B. (1938) The production of homozygous deficient tissues with mutant characteristics by means of the abberrant mitotic behaviour of ring-shaped chromosomes. *Genetics,* 23**,** 315-378.

MEEKER, A. K., HICKS, J. L., GABRIELSON, E., STRAUSS, W. M., MARZO, A. M. D. & ARGANI, P. (2004) Telomere Shortening Occurs in Subsets of Normal Breast Epithelium as well as in Situ and Invasive Carcinoma. *American journal of clinical oncology,* 164**,** 925-935.

MEEKER, A. K., HICKS, J. L., PLATZ, E. A., MARCH, G. E., BENNETT, C. J., DELANNOY, M. J. & MARZO, A. M. D. (2002) Telomere Shortening Is an Early Somatic DNA Alteration in Human Prostate Tumorigenesis Advances in Brief Telomere Shortening Is an Early Somatic DNA Alteration in Human. *Cancer research,* 62**,** 6405-6409.

MEYNE, J., RATLIFF, R. L. & MOYZIS, R. K. (1989) Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 7049-53.

MILLS, R. E., BENNETT, E. A., ISKOW, R. C. & DEVINE, S. E. (2007) Which transposable elements are active in the human genome? *Trends in genetics : TIG,* 23**,** 183-91.

MIN, J., CHOI, E. S., HWANG, K., KIM, J., SAMPATH, S., VENKITARAMAN, A. R. & LEE, H. (2012) The breast cancer susceptibility gene BRCA2 is required for the maintenance of telomere homeostasis. *The Journal of biological chemistry,* 287**,** 5091-101.

MITELMAN, F., MERTENS, F. & JOHANSSON, B. (1997) A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Publishing Group,* special is**,** 417--474.

MIYOSHI, T., KANOH, J., SAITO, M. & ISHIKAWA, F. (2008) Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length. *Science (New York, N.Y.),* 320**,** 1341-4.

MOINFAR, F., MAN, Y. G., BRATTHAUER, G. L., RATSCHEK, M. & TAVASSOLI, F. A. (2000) Genetic abnormalities in mammary ductal intraepithelial neoplasia-flat type ("clinging ductal carcinoma in situ"): a simulator of normal mammary epithelium. *Cancer,* 88**,** 2072-81.

MOKBELL, K., PARRIS, C. N., GHILCHIK, M. & NEWBOLD, R. F. (1999) Telomerase activity in the human breast. *Breast (Edinburgh, Scotland),* 8**,** 208-11.

MOKBELL, K., PARRIS, C. N., GHILCHIK, M. & NEWBOLD, R. F. (1999) Telomerase activity in the human breast. *Breast (Edinburgh, Scotland),* 8**,** 208-11.

MOWAT, E., WILLIAMS, C., JONES, B., MCCHLERY, S. & RAMAGE, G. (2009) The characteristics of Aspergillus fumigatus mycetoma development: is this a biofilm? *Medical mycology : official publication of the International Society for Human and Animal Mycology,* 47 Suppl 1**,** S120-6.

MULLER, H.J. (1938)The remaking of chromosomes. *Collecting Net* 13, 181–198

MUÑOZ-JORDÁN, J. L., CROSS, G. A., DE LANGE, T. & GRIFFITH, J. D. (2001) T-Loops At Trypanosome Telomeres. *The EMBO journal,* 20**,** 579-88.

MURRAY, A. W. & SZOSTAK, J. W. (1986) Construction and behavior of circularly permuted and telocentric chromosomes in Saccharomyces cerevisiae. *Molecular and cellular biology,* 6**,** 3166-72.

NAGELKERKE, A., VAN KUIJK, S. J. A., SWEEP, F. C. G. J., NAGTEGAAL, I. D., HOOGERBRUGGE, N., MARTENS, J. W. M., TIMMERMANS, M. A., VAN LAARHOVEN, H. W. M., BUSSINK, J. & SPAN, P. N. (2011) Constitutive expression of γ-H2AX has prognostic relevance in triple negative breast cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology,* 101**,** 39-45.

NAPOLI, A., VAN DER OOST, J., SENSEN, C. W., CHARLEBOIS, R. L., ROSSI, M. & CIARAMELLA, M. (1999) An Lrp-like protein of the hyperthermophilic archaeon Sulfolobus solfataricus which binds to its own promoter. *Journal of bacteriology,* 181**,** 1474-80.

NAROD, S. A. & FOULKES, W. D. (2004) BRCA1 and BRCA2: 1994 and beyond. *Nature reviews. Cancer,* 4**,** 665-76.

NEGRINI, S., GORGOULIS, V. G. & HALAZONETIS, T. D. (2010) Genomic instability--an evolving hallmark of cancer. *Nature reviews. Molecular cell biology,* 11**,** 220-8.

NEUMANN, A. A. & REDDEL, R. R. (2002) Telomere maintenance and cancer — look , no telomerase. *Nature,* 2**,** 879-885.

NIEDERACHER, D., PICARD, F., VAN ROEYEN, C., AN, H. X., BENDER, H. G. & BECKMANN, M. W. (1997) Patterns of allelic loss on chromosome 17 in sporadic breast carcinomas detected by fluorescent-labeled microsatellite analysis. *Genes, chromosomes & cancer,* 18**,** 181-92.

NISHIZAKI, T., CHEW, K., CHU, L., ISOLA, J., KALLIONIEMI, A., WEIDNER, N. & WALDMAN, F. M. (1997) Genetic alterations in lobular breast cancer by comparative genomic hybridization. *International journal of cancer. Journal international du cancer,* 74**,** 513-7.

NYBERG, K. A., MICHELSON, R. J., PUTNAM, C. W. & WEINERT, T. A. (2002) Toward maintaining the genome: DNA damage and replication checkpoints. *Annual review of genetics,* 36**,** 617-56.

O'DRISCOLL, M. & JEGGO, P. A. (2006) The role of double-strand break repair - insights from human genetics. *Nature reviews. Genetics,* 7**,** 45-54.

OLOVNIKOV, AM (1971) Principle of marginotomy in template synthesis of polynucleotides

Dokl Akad Nauk SSSR. 201(6):1496-9.

OHKI, T. TSURIMOTO, F. ISHIKAWA (2001), In vitro reconstitution of the end replication problem, *Mol. Cell. Biol*. 21 (2001) 5753e5766.

OHTA, M., INOUE, H., COTTICELLI, M.G., KASTURY, K., BAFFA, R., PALAZZO, J., SIPRASHVILI, Z., MORI, M., MCCUE, P., DRUCK, T., (1996). The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell* 84: 587–597.

OKAZAKI, R. SAKABE K, (1966) A unique property of the replicating region of chromosomal DNA. *Biochim Biophys Acta.* Dec 21;129(3):651-4.

OTTAVIANI, A., GILSON, E. & MAGDINIER, F. (2008) Telomeric position effect: from the yeast paradigm to human pathologies? *Biochimie,* 90**,** 93-107.

PAIK, S., SHAK, S., TANG, G., KIM, C., BAKER, J., CRONIN, M., BAEHNER, F. L., WALKER, M. G., WATSON, D., PARK, T., HILLER, W., FISHER, E. R., WICKERHAM, D. L., BRYANT, J. & WOLMARK, N. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine,* 351**,** 2817-26.

PALMIERI, C., CHENG, G. J., SAJI, S., WA, A. & GUSTAFSSON, J.-Å. (2002) Estrogen receptor beta in breast cancer. *Endocrine-related cancer,* 9**,** 1-13.

PANDITA, A., BAYANI, J., PADEROVA, J., MARRANO, P., GRAHAM, C., BARRETT, M., PRASAD, M., ZIELENSKA, M. & SQUIRE, J. A. (2011) Integrated cytogenetic and high-resolution array CGH analysis of genomic alterations associated with MYCN amplification. *Cytogenetic and genome research,* 134**,** 27-39.

PARDUE, M.-L. & DEBARYSHE, P. G. (2003) Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annual review of genetics,* 37**,** 485-511.

PARDUE, M.-L. & DEBARYSHE, P. G. (2003) Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annual review of genetics,* 37**,** 485-511.

PATEL, K. J., YU, V. P. C. C., LEE, H., CORCORAN, A., THISTLETHWAITE, F. C., EVANS, M. J., COLLEDGE, W. H., FRIEDMAN, L. S., PONDER, B. A. J. & VENKITARAMAN, A. R. (1998) Involvement of Brca2 in DNA Repair. *Molecular Cell,* 1**,** 347-357.

PENNANEACH, V., PUTNAM, C. D. & KOLODNER, R. D. (2006) MicroReview Chromosome healing by de novo telomere addition in Saccharomyces cerevisiae. *Molecular microbiology,* 59**,** 1357-1368.

PFEIFFER, V. & LINGNER, J. (2012) TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. *PLoS genetics,* 8**,** e1002747.

PICKETT, H. A., BAIRD, D. M., HOFF-OLSEN, P., MELING, G. I., ROGNUM, T. O., SHAW, J., WEST, K. P. & ROYLE, N. J. (2004) Telomere instability detected in sporadic colon cancers, some showing mutations in a mismatch repair gene. *Oncogene,* 23**,** 3434-43.

PICH, U., FUCHS, J., AND SCHUBERT, I. (1996). How do Alliaceae stabilize their chromosome ends in the absence of TTTAGGG sequences? *Chromosome Res*. 4, 207–213.

PIHET, M., CARRERE, J., CIMON, B., CHABASSE, D., DELHAES, L., SYMOENS, F. & BOUCHARA, J.-P. (2009) Occurrence and relevance of filamentous fungi in respiratory secretions of patients with cystic fibrosis--a review. *Medical mycology : official publication of the International Society for Human and Animal Mycology,* 47**,** 387-97.

POLOTNIANKA, R. M., LI, J. & LUSTIG, A. J. (1998) The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities. *Current biology : CB,* 8**,** 831-4.

POONEPALLI, A., BANERJEE, B., RAMNARAYANAN, K., PALANISAMY, N., PUTTI, T. C. & HANDE, M. P. (2008) Telomere-mediated genomic instability and the clinico-pathological parameters in breast cancer. *Genes, chromosomes & cancer,* 47**,** 1098-109.

PORRO, A., FEUERHAHN, S., REICHENBACH, P. & LINGNER, J. (2010) Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways. *Molecular and cellular biology,* 30**,** 4808-17.

PRESCOTT, J., WENTZENSEN, I. M., SAVAGE, S. A. & DE VIVO, I. (2012) Epidemiologic evidence for a role of telomere dysfunction in cancer etiology. *Mutation research,* 730**,** 75-84.

RAYNAUD, C. M., HERNANDEZ, J., MATHIEU, M.-C., COMMO, F., DELALOGE, S., SABATIER, L. & ANDRE, F. (2010) DNA Damage Repair and Telomere Length in Normal Breast , Preneoplastic Lesions , and Invasive Cancer. *American journal of clinical oncology,* 33**,** 341-345.

REDON, S., REICHENBACH, P. & LINGNER, J. (2010) The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic acids research,* 38**,** 5797-5806.

REDON, S., REICHENBACH, P. & LINGNER, J. (2010) The non-coding RNA TERRA is a

natural ligand and direct inhibitor of human telomerase. *Nucleic acids research,* 38**,** 5797-806.

RICHARDS, E. J. & AUSUBEL, F. M. (1988) Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. *Cell,* 53**,** 127-36.

RICHARDSON, C., MOYNAHAN, M. E. & JASIN, M. (1998) Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes & development,* 12**,** 3831-42.

RICHIE, R. C. & SWANSON, J. O. (2003) Breast Cancer : A Review of the Literature. *Journal of insurance medicine (New York, N.Y.),* 35**,** 85-101.

RIETHMAN, H. (2008) Human telomere structure and biology. *Annual review of genomics and human genetics,* 9**,** 1-19.

RIETHMAN, H., AMBROSINI, A., CASTANEDA, C., FINKLESTEIN, J., HU, X.-L., MUDUNURI, U., PAUL, S. & WEI, J. (2004) Mapping and Initial Analysis of Human Subtelomeric Sequence Assemblies. *Genome research,* 14**,** 18-28.

RIETHMAN, H., AMBROSINI, A., CASTANEDA, C., FINKLESTEIN, J., HU, X.-L., MUDUNURI, U., PAUL, S. & WEI, J. (2004) Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome research,* 14**,** 18-28.

RIETHMAN, H., AMBROSINI, A. & PAUL, S. (2005) Human subtelomere structure and variation. *Chromosome Research,* 13**,** 505-515.

ROGAN, E. M., BRYAN, T. M., HUKKU, B., MACLEAN, K., CHANG, A. C., MOY, E. L., ENGLEZOU, A., WARNEFORD, S. G., DALLA-POZZA, L. & REDDEL, R. R. (1995) Alterations in p53 and p16 INK4 Expression and Telomere Length during Spontaneous Immortalization of Li-Fraumeni Syndrome Fibroblasts. *Mol Cell.,* 15**,** 4745-4753.

ROLLI, V. SCHREIBER.,(1999) "PARP-2, a novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase," *Journal of Biological Chemistry*, vol. 274, no. 25, pp. 17860–17868,

ROOS G, KROBER A, GRABOWSKI P (2007). Short telomeres are associated with genetic complexity, high risk genomic aberrations, and short survival in chronic lymphocytic leukemia. *Blood.;*111(4):2246-2252.

ROY, R., CHUN, J. & POWELL, S. N. (2012) BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature reviews. Cancer,* 12**,** 68-78.

RUDD, M. K. (2007) Subtelomeres : Evolution in the Human Genome. *Encyclopedia of life sciences***,** 1-9.

RUDD, M. K., FRIEDMAN, C., PARGHI, S. S., LINARDOPOULOU, E. V., HSU, L. & TRASK, B. J. (2007) Elevated rates of sister chromatid exchange at chromosome ends. *PLoS genetics,* 3**,** e32.

RUDOLPH, K. L., MILLARD, M., BOSENBERG, M. W. & DEPINHO, R. A. (2001) Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. *Nature genetics,* 28**,** 155-9.

SAINSBURY, J. R. C., ANDERSON, T. J. & MORGAN, D. A. L. (2000) Breast cancer Classification of invasive breast cancers. *Clinical review,* 321**,** 745-751.

SAN FILIPPO, J., SUNG, P. & KLEIN, H. (2008) Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry,* 77**,** 229-57.

SANDELL, L. L., GOTTSCHLING, D. E. & ZAKIAN, V. A. (1994) Transcription of a yeast telomere alleviates telomere position effect without affecting chromosome stability. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 12061-5.

SARETZKI, G., SITTE, N., MERKEL, U., WURM, R. E. & ZGLINICKI, T. V. (1999) Telomere shortening triggers a p53-dependent cell cycle arrest via accumulation of G-rich single stranded DNA fragments. *Oncogene,* 18**,** 5148-5158.

SCHRÖDER, C. P., WISMAN, G. B., DE JONG, S., VAN DER GRAAF, W. T., RUITERS, M. H., MULDER, N. H., DE LEIJ, L. F., VAN DER ZEE, A. G. & DE VRIES, E. G. (2001) Telomere length in breast cancer patients before and after chemotherapy with or without stem cell transplantation. *British journal of cancer,* 84**,** 1348-53.

SCHULZ, W. A. (2006) L1 retrotransposons in human cancers. *Journal of biomedicine & biotechnology,* 2006**,** 83672.

SCULLY, R., CHEN, J., OCHS, R. L., KEEGAN, K., HOEKSTRA, M., FEUNTEUN, J. & LIVINGSTON, D. M. (1997) Dynamic Changes of BRCA1 Subnuclear Location and Phosphorylation State Are Initiated by DNA Damage. *Cell,* 90**,** 425-435.

SEN, S. (2000) Aneuploidy and cancer. *Current opinion in oncology,* 12**,** 82-88.

SFEIR, A., KABIR, S., VAN OVERBEEK, M., CELLI, G. B. & DE LANGE, T. (2010) Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science (New York, N.Y.),* 327**,** 1657-61.

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

SHAY, J. W. & WRIGHT, W. E. (2001) Telomeres and Telomerase : Implications for Cancer and Aging. *Radiation Research,* 155**,** 188-193.

SHAY, J. W., ZOU, Y., HIYAMA, E. & WRIGHT, W. E. (2001) Telomerase and cancer. *Human*

molecular genetics, 10**,** 677-686.

SHEN, J., GAMMON, M. D., TERRY, M. B., BRADSHAW, P. T., WANG, Q., TEITELBAUM, S. L., NEUGUT, A. I. & SANTELLA, R. M. (2012) Genetic polymorphisms in telomere pathway genes, telomere length, and breast cancer survival. *Breast cancer research and treatment,* 134**,** 393-400.

SHEN, J., TERRY, M. B., GURVICH, I., LIAO, Y., SENIE, R. T. & SANTELLA, R. M. (2007) Short Telomere Length and Breast Cancer Risk : A Study in Sister Sets in Sister Sets. *Cancer research,* 67**,** 5538-5544.

SHEN, M., HAGGBLOM, C., VOGT, M., HUNTER, T. & LU, K. P. (1997) Characterization and cell cycle regulation of the related human telomeric proteins Pin2 and TRF1 suggest a role in mitosis. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 13618-23.

SHEN, Z. (2011) Genomic instability and cancer : an introduction. *Journal of molecular cell biology,* 3**,** 1-3.

SHILOH, Y. (2006) The ATM-mediated DNA-damage response: taking shape. *Trends in biochemical sciences,* 31**,** 402-10.

SHRIVASTAV, M., HARO, L. P. D. & NICKOLOFF, J. A. (2008) Regulation of DNA double-strand break repair pathway choice. 18**,** 134-147.

SMITH, S. (1998) Tankyrase, a Poly(ADP-Ribose) Polymerase at Human Telomeres. *Science,* 282**,** 1484-1487.

SMOGORZEWSKA, A., KARLSEDER, J., HOLTGREVE-GREZ, H., JAUCH, A. & LANGE, T. D. (2002) DNA Ligase IV-Dependent NHEJ of Deprotected Mammalian Telomeres in G1 and G2. *Current biology : CB,* 12**,** 1635-1644.

STANSEL, R. M., DE LANGE, T. & GRIFFITH, J. D. (2001) T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang. *The EMBO journal,* 20**,** 5532-40.

STEWART, S. A. & WEINBERG, R. A. (2006) Telomeres: cancer to human aging. *Annual review of cell and developmental biology,* 22**,** 531-57.

SUBHAWONG, A. P., HEAPHY, C. M., ARGANI, P., KONISHI, Y., KOUPRINA, N., NASSAR, H., VANG, R. & MEEKER, A. K. (2009) The alternative lengthening of telomeres phenotype in breast carcinoma is associated with HER-2 overexpression. *Modern Pathology,* 22**,** 1423-1431.

SUWANTO, A. & KAPLAN, S. (1989) Physical and genetic mapping of the Rhodobacter sphaeroides 2.4.1 genome: presence of two unique circular chromosomes. *Journal of bacteriology,* 171**,** 5850-9.

SVENSON, U., NORDFJÄLL, K., STEGMAYR, B., MANJER, J., NILSSON, P., TAVELIN, B., HENRIKSSON, R., LENNER, P. & ROOS, G. (2008) Breast cancer survival is associated with telomere length in peripheral blood cells. *Cancer research,* 68**,** 3618-23.

SÝKOROVÁ, E., LIM, K. Y., KUNICKÁ, Z., CHASE, M. W., BENNETT, M. D., FAJKUS, J. & LEITCH, A. R. (2003) Telomere variability in the monocotyledonous plant order Asparagales. *Proceedings. Biological sciences / The Royal Society,* 270**,** 1893-904.

SYMINGTON, L. S. & GAUTIER, J. (2011) Double-Strand Break End Resection and Repair Pathway Choice. *Annual review of genetics,* 45**,** 247-273.

SZOSTAK, J. W., ORR-WEAVER, T. L., ROTHSTEIN, R. J. & STAHL, F. W. (1983) The Double-Strand-Break Repair Model for Recombination. *Cell,* 33**,** 25-35.

TANAKA, H., ABE, S., HUDA, N., TU, L., BEAM, M. J., GRIMES, B. & GILLEY, D. (2012) Telomere fusions in early human breast carcinoma. *PNAS,* 109**,** 14098-14103.

TANKIMANOVA, M., CAPPER, R., LETSOLO, B. T., ROWSON, J., JONES, R. E., BRITT-COMPTON, B., TAYLOR, A. M. R. & BAIRD, D. M. (2012) Mre11 modulates the fidelity of fusion between short telomeres in human cells. *Nucleic acids research,* 40**,** 2518-26.

TARSOUNAS, M., MUÑOZ, P., CLAAS, A., SMIRALDO, P. G., PITTMAN, D. L., BLASCO, M. A. & WEST, S. C. (2004) Telomere maintenance requires the RAD51D recombination/repair protein. *Cell,* 117**,** 337-47.

TAYLOR, B. S., BARRETINA, J., SOCCI, N. D., DECAROLIS, P., LADANYI, M., MEYERSON, M., SINGER, S. & SANDER, C. (2008) Functional copy-number alterations in cancer. *PloS one,* 3**,** e3179.

TEJERA, A. M., STAGNO D'ALCONTRES, M., THANASOULA, M., MARION, R. M., MARTINEZ, P., LIAO, C., FLORES, J. M., TARSOUNAS, M. & BLASCO, M. A. (2010) TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. *Developmental cell,* 18**,** 775-89.

TELLI, M & FORD, J (2010) PARP inhibitors in breast cancer. *Clin Adv Hematol Oncol*. Sep;8(9):629-35.

THIAGALINGAM, S., LAKEN, S., WILLSON, J. K., MARKOWITZ, S. D., KINZLER, K. W., VOGELSTEIN, B. & LENGAUER, C. (2001) Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 2698-702.

THORSLUND, T., MCILWRAITH, M. J., COMPTON, S. A., LEKOMTSEV, S.,

PETRONCZKI, M., GRIFFITH, J. D. & WEST, S. C. (2010) The breast cancer tumor suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA. *Nature structural & molecular biology,* 17**,** 1263-5.

TIBBETTS, R. S., CORTEZ, D., BRUMBAUGH, K. M., TIBBETTS, R. S., CORTEZ, D., BRUMBAUGH, K. M., SCULLY, R., LIVINGSTON, D., ELLEDGE, S. J. & ABRAHAM, R. T. (2000) Functional interactions between BRCA1 and the checkpoint kinase ATR during genotoxic stress. *Genes & development,* 14**,** 2989-3002.

TONG, W. M., CORTES, U. & WANG, Z. Q. (2001) Poly(ADP-ribose) polymerase: a guardian angel protecting the genome and suppressing tumorigenesis. *Biochimica et biophysica acta,* 1552**,** 27-37.

TOUSSAINT O, DUMONT P, REMACLE J, DIERICK JF, PASCAL T, FRIPPIAT C, MAGALHAES JP, ZDANOV S, CHAINIAUX F. (2002) Stress-induced premature senescence or stress-induced senescence-like phenotype: one in vivo reality, two possible definitions? *ScientificWorldJournal*. Jan 29;2:230-47

UREN, A. G., MIKKERS, H., KOOL, J., VAN DER WEYDEN, L., LUND, A. H., WILSON, C. H., RANCE, R., JONKERS, J., VAN LOHUIZEN, M., BERNS, A. & ADAMS, D. J. (2009) A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nature protocols,* 4**,** 789-98.

UROSKIE, T. W. & COLEN, L. B. (2004) History of breast reconstruction. *Seminars in plastic surgery,* 18**,** 65-9.

VAN DER KLIFT, H., WIJNEN, J., WAGNER, A., VERKUILEN, P., TOPS, C., OTWAY, R., KOHONEN-CORISH, M., VASEN, H., OLIANI, C., BARANA, D., MOLLER, P., DELOZIER-BLANCHET, C., HUTTER, P., FOULKES, W., LYNCH, H., BURN, J., MÖSLEIN, G. & FODDE, R. (2005) Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). *Genes, chromosomes & cancer,* 44**,** 123-38.

VAN ’T VEER LJ, DAI H, VAN DE VIJVER MJ, (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature*;415:530-6.

VAN DEVIJVERMJ, HEYD, VAN’TVEER LJ, (2002). A geneexpression signature as a predictor of survival in breast cancer. *NEngl JMed* ;347:1999-2009.

VAN STEENSEL, B. A. D. L., T (1997) Control of telomere length by the human telomeric protein TRF1. *Nature,* 385**,** 740-743.

VAN STEENSEL, B., SMOGORZEWSKA, A., DE LANGE, T.(1998) TRF2 protects human telomeres from end-to-end fusions. *Cell* 92:401–413

VAZIRI, H. & BENCHIMOL, S. (1998) Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. *Current biology : CB,* 8**,** 279-282.

VENDITTI, S., VEGA-PALAS, M. A. & DI, E. (1999) Heterochromatin Organization of a Natural Yeast Telomere. 274**,** 1928-1933.

VENTURA, A., KIRSCH, D. G., MCLAUGHLIN, M. E., TUVESON, D. A., GRIMM, J., LINTAULT, L., NEWMAN, J., RECZEK, E. E., WEISSLEDER, R. & JACKS, T. (2007) Restoration of p53 function leads to tumour regression in vivo. *Nature,* 445**,** 661-5.

VIJVER, M., HE, Y., VAN'T VEER, L., DAI, H., HART, A., VOSKUIL, D., SCHREIBER, G., PETERSE, K., GLAS, A., DELHAYE, L., VAN DER VELDE, T., BARTELINK, H. & BERNARDS, R. (2002) A gene-expression signature is a predictor of survival in breast cancer. *The New England journal of medicine,* 347**,** 1999-2009.

VOLPICELLA, M., LEONI, C., COSTANZA, A., FANIZZA, I., PLACIDO, A. & CECI, L. R. (2012) Genome Walking by Next Generation Sequencing Approaches. *Biology,* 1**,** 495-507.

VULLIAMY, T., MARRONE, A., GOLDMAN, F., DEARLOVE, A., BESSLER, M., MASON, P. J. & DOKAL, I. (2001) The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Letters to nature,* 413.

VULLIAMY, T., MARRONE, A., SZYDLO, R., WALNE, A., MASON, P. J. & DOKAL, I. (2004) Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC. *Nature genetics,* 36**,** 447-9.

VULLIAMY, T. J. & DOKAL, I. (2008) Dyskeratosis congenita : The diverse clinical presentation of mutations in the telomerase complex. *Biochimie,* 90**,** 122-130.

WALDMAN, A. S. & WALDMAN, B. C. (1991) Stimulation of intrachromosomal homologous recombination in mammalian cells by an inhibitor of poly(ADP-ribosylation). *Nucleic acids research,* 19**,** 5943-7.

WALKER, J. R., CORPINA, R. A. & GOLDBERG, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature***,** 607-614.

WATSON, J.D. (1972).Origin of concatameric T7 DNA. *Nat. New Biol*. 239, 197–201

WATSON, J.D. & CRICK, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature 171*, 737–738.

WEIGELT, B., PETERSE, J. L. & VAN 'T VEER, L. J. (2005) Breast cancer metastasis: markers and models. *Nature reviews. Cancer,* 5**,** 591-602.

WELSCH, C. W. & NAGASAWA, H. (1977) Prolactin and Murine Mammary Tumorigenesis : A Review Prolactin and Murine Mammary Tumorigenesis : A Review. *Cancer research,* 37**,** 951-963.

WESLEY, D. (2007) Secondary cancer rates following breast cancer diagnosis. *Journal of insurance medicine (New York, N.Y.),* 39**,** 98-106.

WESSELS, L. F. A., WELSEM, T. V. & HART, A. A. M. (2002) Molecular Classification of Breast Carcinomas by Comparative Genomic Hybridization : a Specific Somatic Genetic Profile for BRCA1 Tumors Molecular Classification of Breast Carcinomas by Comparative Genomic Hybridization : a Specific Somatic Genetic Profil. *Cancer research,* 62**,** 7110-7117.

WHYTE, J., BERGIN, O., BIANCHI, A., MCNALLY, S. & MARTIN, F. (2009) Key signalling nodes in mammary gland development and cancer. Mitogen-activated protein kinase signalling in experimental models of breast cancer progression and in mammary gland development. *Breast cancer research : BCR,* 11**,** 209.

WILKIE, A., LAMB, J., HARRIS, P., FINNEY, R. & HIGGS, D. (1990) A truncated human chromosome 16 associated with alpha thalassemia is stabilized by the addtion of telomeric repeat (TTAGGG)n. *Nature,* 346**,** 868.

WONG, G. , TIWARI, R., BORGEN, P., CORDEN-CARDO, C., OSBORNE, M.(1992) HER-2/neu amplification and overexpression in primary human breast cancer is associated with early metastasis. *Anticancer Research.* Mar-Apr;12(2):419-25.

WOOSTER, R., ALVES, A. M., CARR, F. J., NEUHAUSEN, S. L., MANGION, J., QUIRK, Y., FORD, D., COLLINS, N., NGUYEN, K., SEAL, S., TRAN, T., AVERILL, D., FIELDS, P., MARSHALL, G., NAROD, S., LENOIR, G. M., LYNCH, H., FEUNTEUN, J., DEVILEE, P., CORNELISSE, C. J., MENKO, F. H., DALY, P. A., ORMISTON, W., MCMANUS, R., PYE, C., LEWIS, C. M., CANNON-ALBRIGHT, L. A., PETO, J., PONDER, B. A. J., SKOLNICK, M. H., EASTON, D. F., GOLDGAR, D. E. & STRATTON, M. R. (1994) Localization of a Breast Cancer Susceptibility Gene , BRCA2 , to Chromosome 13q1 2-13. *Science,* 265**,** 15-17.

WRIGHT, W., SHAY, J (1992) The two-stage mechanism controlling cellular senescence and

immortalization. *Experimental Gerontology*. Jul-Aug;27(4):383-9.

WRIGHT, W. E., TESMER, V. M., HUFFMAN, K. E., LEVENE, S. D. & SHAY, J. W. (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes & Development,* 11**,** 2801-2809.

WYMAN, C. & KANAAR, R. (2006) DNA Double-Strand Break Repair : All ' s Well that Ends Well. *Annual review of genetics,* 40**,** 363-383.

WYMAN, C. & KANAAR, R. (2006) DNA double-strand break repair: all's well that ends well. *Annual review of genetics,* 40**,** 363-83.

XIN, H., LIU, D. & SONGYANG, Z. (2008) The telosome/shelterin complex and its functions. *Genome biology,* 9**,** 232.

XU, L. A. B., ELIZABETH (2007) Human cancer cells harbor T-stumps, a distinct class of extremely short telomeres. *Mol Cell.,* 28**,** 315-327.

YANEZ, G. H., KHAN, S. J., LOCOVEI, A. M., PEDROSO, I. M. & FLETCHER, T. M. (2005) DNA structure-dependent recruitment of telomeric proteins to single-stranded/double-stranded DNA junctions. *Biochemical and boiphysical research communications BBRC,* 328**,** 49-56.

YELAMOS, J., FARRES, J., LLACUNA, L., AMPURDANES, C. & MARTIN-CABALLERO, J. (2011) PARP-1 and PARP-2: New players in tumour development. *American journal of cancer research,* 1**,** 328-346.

YOO, S. & DYNAN, W. S. (1999) Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic acids research,* 27**,** 4679-86.

YOUNG, N. S. (2010) Telomere biology and telomere diseases: implications for practice and research. *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program,* 2010**,** 30-5.

YU, G.-L. & BLACKBURN, E. (1991) Developmentally Programmed Chromosomes by Telomerase Healing of in Tetrahymena. *Cell,* 67**,** 623-632.

YU, V. P. C. C., KOEHLER, M., STEINLEIN, C., SCHMID, M., HANAKAHI, L. A., GOOL, A. J. V., WEST, S. C. & VENKITARAMAN, A. R. (2000) Gross chromosomal rearrangements and genetic exchange between nonhomologous chromosomes following BRCA2 inactivation. *Genes & development,* 14**,** 1400-1406.

ZHANG, J., WILLERS, H., FENG, Z., JAGADISH, C., KIM, S., WEAVER, D. T., CHUNG, J. H., SIMON, N., XIA, F., ZHANG, J., WILLERS, H., FENG, Z., GHOSH, J. C., KIM, S., WEAVER, D. T., CHUNG, J. H., POWELL, S. N. & XIA, F. (2004) Chk2 Phosphorylation of BRCA1 Regulates DNA Double-Strand Break Repair Chk2 Phosphorylation of BRCA1 Regulates DNA Double-Strand Break Repair. *Molecular and cellular biology,* 24**,** 708.

ZHENG, Y.-L., LOFFREDO, C. A., SHIELDS, P. G. & SELIM, S. M. (2009) Chromosome 9 arm-specific telomere length and breast cancer risk. *Carcinogenesis,* 30**,** 1380-6.

ZHONG, Z., SHIUE, L., KAPLAN, S. & DE LANGE, T. (1992) A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Molecular and cellular biology,* 12**,** 4834-43.

ZHOU, X., MEEKER, A. K., MAKAMBI, K. H., KOSTI, O., KALLAKURY, B. V. S., SIDAWY, M. K., LOFFREDO, C. A. & Ã, Y.-L. Z. (2012) Telomere length variation in normal epithelial cells adjacent to tumor : potential biomarker for breast cancer local recurrence. *Carcinogenesis,* 33**,** 113-118.

ZHU, X.-D., NIEDERNHOFER, L., KUSTER, B., MANN, M., HOEIJMAKERS, J. H. J. & DE LANGE, T. (2003) ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. *Molecular cell,* 12**,** 1489-98.

ZOU, L. & ELLEDGE, S. J. (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science (New York, N.Y.),* 300**,** 1542-8.

ZOU, Y., SFEIR, A., GRYAZNOV, S. M., SHAY, J. W. & WRIGHT, W. E. (2004) Does a Sentinel or a Subset of Short Telomeres Determine Replicative Senescence ? 15**,** 3709-3718.

Appendix 1b

wcbnum	Age at Operation	pTNM	NPI Score	Breast Grade	ER Score	pgr score	her2score	Her2 2+ : Pos or neg following FISH	XpYp tel
RR6BL0000011	84		0	III	Negative				1.075
RVCC40000033	75	T2 N3a Mx	6.82	III	Negative	Negative	3	positive	1.290
RVCC40000753	68	T1c No Mx	4.36	III	positive		0	negative	1.664
RR6BL0000004	82	T2 No Mx	0	II	positive				1.780
RR6BL0000120	69		0	II					1.818
jRVCC40000076	43	T2 N2a Mx	6.46	III	Negative	Negative	0	negative	1.845
RR6BL0000001	72	T2 N1 Mx	0	III	positive	Negative	0	negative	2.021
RVCC40000022	79	T2N1miMx	5.42	III	Negative	Negative	3	positive	2.182
RR6BL0000008	61	T1 No Mx	0	II	positive				2.290
RVCC40001036	73	T2N1Mx	5.84	III	Negative	Negative	2	negative	2.328
RR6BL0000350	59		0	II	positive	positive	1	negative	2.458
RR6BL0000083	75	T2 N1 Mx	0	III	positive				2.512
RR6BL0000274	79	T1 No	0	I	positive		0	negative	2.517
RVCC40000124	45	T2N1aMx	4.8	II	positive	positive	0	negative	2.617
RVCC40000622	59	T2N0Mx	4.44	III	positive	positive	0	negative	2.660
RVCC40000965	73	T2NOMX	4.72	III	Negative	Negative	3	positive	2.669
RR6BL0000013	48	T1 No Mx	0	III	Negative	Negative	2	negative	2.680
RVCC40000021	63	T1c No Mx	3.28	II	positive		0	negative	2.686
RR6BL0000039	34	T2 No Mx	0	III	Negative	positive			2.700
RVCC40000665	65		0	II	positive		0	negative	2.701
RVCC40000072	65	T1c No Mx	4.36	III	Negative	positive	0	negative	2.729
RR6BL0000045	56	T2 No Mx	0	II	positive				2.730
RVCC40000907	53	T2mN1Mx	4.54	II	positive		0	negative	2.793
RR6BL0000067	58	T2 No Mx	0	III	positive	Negative	2	negative	2.800
RR6BL0000072	58	T2 No Mx	0	III	Negative	Negative	3	positive	2.920
RVCC40000135	83	T1cNoMx	2.4	I	positive				2.946
RVCC40000009	68	T2 N2a Mx	6.6	III	positive		3	positive	2.960
RVCC40000025	38		4.3	III	positive		3	positive	2.980
RVCC40000030	74	T1cNoMx	3.34	II	positive				3.040
RR6BL0000386	75	t1n0mx	0	II	positive	positive	2	negative	3.044

Appendix 1b

RR6BL0000002	73	T2 N1 Mx	0	II	Negative	Negative			3.118
RVCC40000752	47	T2N0Mx	3.4	II	positive		1	negative	3.153
RR6BL0000020	86		0	II	positive				3.158
RVCC40001469	81	T2NoMx	2.6	I	positive	positive	1	negative	3.225
RVCC40000042	85	T2 No Mx	3.5	II	positive				3.230
RVCC40000144	39	T2N0Mx	4.6	III	Negative	Negative	1	negative	3.260
RVCC40000051	42	T1c No Mx	2.24	I	positive				3.260
RR6BL0000171	45	T2 N1	0	II	positive		1	negative	3.265
RVCC40001055	50	T1cNoMx	3.28	II	positive		1	negative	3.353
RVCC40000696	42	T2N3aMx	6.46	III	Negative	Negative	3	positive	3.366
RVCC40000089	66	T1c N1 Mx	5.26	III	positive	positive	0	negative	3.434
RVCC40000155	78	T2 N1 Mx	5.48	III	positive				3.450
RVCC40000165	47	T2 N1m Mx	5.6	III	positive		0	negative	3.500
RVCC40000795	65	T2N1aMx	3.56	I	positive		1	negative	3.520
RVCC40000070	73	T2 No Mx	3.56	II	Negative	Negative	1	negative	3.542
RVCC40000029	58	T3 N1 Mx	5.02	II	positive	positive	0	negative	3.554
RR6BL0000104	59	T2 No Mx	0	III	positive	Negative	2	negative	3.580
RVCC40000631	79	T2N0Mx	4.66	III	positive		3	positive	3.602
RVCC40000166	69	T1c No Mx	4.34	III	Negative		3	positive	3.700
RVCC40000936	61	T3N1miMx	5.08	II	positive		0	negative	3.702
RR6BL0000111	79	T1 N1 Mx	0	II	positive	Negative			3.719
RVCC40000035	68	T2 N1a Mx	5.46	III	positive	positive	0	negative	3.720
RR6BL0000232	36		0	II	positive	Negative	3	positive	3.721
RVCC40000016	60	T2 N12 Mx	4.54	II	positive	positive	2	negative	3.748
RR6BL0000204	69		0	II	positive	positive	0	negative	3.763
RVCC40000026	45	T1c No Mo	2.28	I	positive	positive	0	negative	3.778
RVCC40000633	68	T2aNoMx	3.44	II	positive	positive	2	negative	3.808
RR6BL0000252	52		0	III	positive	positive	1	negative	3.822
RVCC40000686	49	T2N1miMx	5.52	III	positive		0	negative	3.830
RVCC40000039	46	pT2 No Mx	4.6	III	Negative	Negative	0	negative	3.833
RVCC40000063	81	T3 N1 Mx	6.3	III	Negative	Negative	1	negative	3.879
RVCC40000642	70	T4bNoMx	3.28	II	positive	positive	0	negative	3.892
RR6BL0000557	83	T2N0MX	0	III	positive	positive	3	positive	3.961

Appendix 1b

RR6BL0000070	80	T2 N2 Mx	0	III	Negative	Negative	3	positive	3.977
RVCC40001236	55	T1cN0Mx	4.36	III	positive	positive	0	negative	3.981
RR6BL0000623	48	t2n0mx	0	III	positive	Negative	3	positive	3.987
RR6BL0000214	68		0	III	positive	positive	0	negative	4.018
RVCC40000015	87	T2NoMx	4.66	III	positive	Negative	3	positive	4.059
RVCC40000539	70	T1cN0Mx	3.3	II	positive	positive	0	negative	4.080
RVCC40000134	70	T2 No Mx	4.42	III	positive	Negative	0	negative	4.080
RR6BL0000005	53	T2 N1 Mx	0	II	positive		1	negative	4.092
RVCC40000573	46	T2NoMx	3.96	I	positive	positive	0	negative	4.100
RVCC40000011	58	T2 N1a Mx	4.44	II	positive		2	negative	4.110
RVCC40000001	71	T2 No Mx	4.44	III	Negative	Negative	3	positive	4.111
RR6BL0000198	51	T2 N2 Mx	0	III	positive		3	positive	4.120
RVCC40001111	48	T2N0Mx	4.44	III	positive	positive	1	negative	4.124
RR6BL0000186	80		0	I	positive	positive	1	negative	4.146
RR6BL0000019	49	T2 No Mx	0	I	positive		1	negative	4.163
RR6BL0000528	56	T1N0MX	0	II	positive	Negative	3	positive	4.170
RVCC40000638	91	T2N1	0	II	positive	positive	0	negative	4.200
RVCC40000583	47	T2N2aMx	5.98	II	positive	positive	2	negative	4.200
RVCC40000559	61	T1cN1aMx	4.24	II	positive	positive	2	negative	4.219
RVCC40000589	76	T2mN3aMx	6.8	III	positive	positive	0	negative	4.220
RVCC40000632	33	T2N1Mx	5.76	III	Negative	Negative	0	negative	4.230
RR6BL0000117	67	T2 N1 Mx	0	II	positive		1	negative	4.290
RR6BL0000141	48		0	III	Negative		3	positive	4.378
RR6BL0000213	58	T1 No	0	II	positive		3	positive	4.459
RR6BL0000033	84	T2 N1 Mx	0	II	positive				4.468
RVCC40000647	65	T2N0Mx	4.44	III	Negative	Negative	2	negative	4.500
RR6BL0000245	86		0	II	positive	positive	0	negative	4.545
RR6BL0000064	40	T1 No Mx	0	II	positive	positive	1	negative	4.635
RVCC40000031	54	T1c No Mx	4.32	III	Negative	Negative	1	negative	4.659
RR6BL0000184	42	T3N0MX	0	III	Negative	Negative	0	negative	4.659
RVCC40000687	49	T2 No Mx	3.44	II	positive		1	negative	4.690
RVCC40000649	60	T2N1aMx	5.62	III	Negative	Negative	1	negative	4.700
RVCC40000597	48	T2N1aMx	4.6	II	positive	positive	0	negative	4.701

RVCC40000067	48	T1 N1 Mo	5.34	III	positive	positive	1	negative	4.866
RVCC40000591	50	T2N1Mx	5.66	III	positive	positive	1	negative	4.871
RR6BL0000199	70	T1 No Mx	0	II	positive	positive	1	negative	4.894
RR6BL0000535	46	T2N0MX	0	III	positive	positive	3	positive	4.895
RR6BL0000235	53	T2	0	III	positive	Negative	3	positive	4.896
RR6BL0000205	72	Unknown	0	II	positive	positive	0	negative	4.930
RVCC40000786	60	T2NoMx	3.44	II	positive				4.991
RVCC40000776	65	T1cN0Mx	3.24	II	positive		0	negative	5.138
RR6BL0000131	69		0	II	positive	positive	1	negative	5.146
RR6BL0000409	43	t2n1m0	0	III	positive	positive	3	positive	5.151
RVCC40000348	87	T2NoMx	2.44	I	positive	positive	0	negative	5.203
RVCC40000698	70	T2N0Mx	4.82	III	Negative	Negative	0	negative	5.300
RR6BL0000230	43		0	II	positive	positive	1	negative	5.345
RVCC40000555	76	T2 N3a Mx	6.64	III	positive	Negative	0	negative	5.421
RVCC40000643	42	T1N0Mx	3.26	II	positive		2	positive	5.463
RR6BL0000202	66		0	II	positive		0	negative	5.479
RVCC40000090	84	T1c No Mx	3.4	II	positive				5.519
RR6BL0000571	42	T2N1biMO	0	III	positive	positive	3	positive	5.578
RVCC40000755	73	T2N0Mx	4.28	III	positive		2	negative	5.600
RR6BL0000540	44	T1cN1biMX	0	III	positive	Negative	3	positive	5.642
RVCC40000201	66	T2 N1 Mx	5.7	III	Borderline	Borderline	3	positive	5.736
RVCC40000154	49	T2 N1 Mx	6	III	Negative	Negative	2	negative	5.750
RVCC40000758	52	T2N3aMx	6.58	III	positive		0	negative	5.800
RR6BL0000134	68		0	II	positive	positive	0	negative	5.812
RR6BL0000071	71	T2 N1 Mx	0	II	Negative	Negative	0	negative	5.847
RR6BL0000259	75		0	III	positive	Negative	3	positive	6.130
RVCC40000737	78	T4bNoMx	4.5	III	positive	Borderline	2	negative	7.036
RVCC40001060	45	T3N2Mx	0	III	Negative	Negative	3	positive	7.046
RR6BL0000225	67	Unknown	0	I	positive	positive	0	negative	7.268
RVCC40000071	85	T3 N1	6.6	III	positive	Negative			7.852
RR6BL0000130	47	T1 No Mx	0	III	Negative	Negative	2	negative	8.460
RVCC40000020	55	T2 No Mx	4.44	III	Negative	Negative	3	positive	9.975

Summary

Mean Age	ER +	PR +	ER/PR +	HER2 +
62	94	2	40	29

Appendix 1b: A summary of clinical data that was used in subsequent analysis is represented alongside patient number (wcbnum). Details include TNM score; NPI score; Grade given to each patient; ER/PR/HER2 status and age at operation. Telomere length means used for Kaplan meier analysis at XpYp are shown alongside. A summary of patient clinical features shows mean age and the number of patients with ER+/PR+ and HER+ breast cancer

Appendix 1b

Appendix 1a Breast (IDC) follow up summary

WCB Num	sex	age	Grade	Total Nodes	Positive Nodes	Time (days)	Alive	XpYp Mean	XpYp SD	17p Mean	17pSD
RR6BL0000005	F	53	II	5	5	1725	Yes	3.016849398	1.96	5.326209	2.32
RR6BL0000019	F	49	I	10	0	2001	Yes	2.701871795	3.44	4.23181	1.98
RR6BL0000020	F	86	II	11	0	2432	Yes	5.436451613	2.36	4.797212	1.51
RR6BL0000033	F	84	II	6	3	1658	Yes	1.29	1.53	4.991378	1.88
RR6BL0000064	F	40	II	10	0	2271	No	1.79115625	2.73	3.161435	1.46
RR6BL0000070	F	80	III	7	5	1382	Yes	4.518212121	4.01	3.96927	2.14
RR6BL0000071	F	71	II	18	3	504	Yes	3.750581818	2.21	5.420231	2.13
RR6BL0000083	F	75	III	11	7	1900	Yes	9.947666667	2.25	8.2615	4.21
RR6BL0000104	F	59	III	24	0	1773	No	3.57508	2.55	3.861723	2.20
RR6BL0000111	F	79	II	14	0	712	No	6.102722543	3.53	5.932856	3.05
RR6BL0000117	F	67	II	9	0	1808	No	3.959614815	2.38	3.987069	1.23
RR6BL0000120	F	69	II	4	2	803	Yes	4.607795322	2.93	4.817007	2.85
RR6BL0000130	F	47	III	18	0	1780	Yes	3.991190476	2.41	5.218654	3.22
RR6BL0000131	F	69	II	4	3	1766	Yes	4.350948052	2.81	4.591596	2.34
RR6BL0000134	F	68	II	13	0	1757	Yes	3.675360544	2.18	4.169907	1.95
RR6BL0000141	F	48	III	14	12	362	Yes	5.118796117	2.29	5.512888	2.46
RR6BL0000170	F	66	I	5	0	1694	Yes	4.838508671	3.10	5.341982	2.45
RR6BL0000171	F	45	II	10	5	1694	No	5.723454545	3.70	6.391345	2.80
RR6BL0000184	F	42	III	15	4	794	No	4.497176923	3.70	4.358244	1.88
RR6BL0000186	F	80	I	2	2	2147	Yes	4.963623529	2.70	4.603157	2.18
RR6BL0000198	M	51	III	7	6	882	Yes	3.954214815	2.50	6.282703	2.96
RR6BL0000199	F	70	II	13	0	1729	Yes	5.123664286	2.62	4.784091	1.89
RR6BL0000202	F	66	II	4	2	1884	Yes	4.096993289	2.73	4.454908	2.55
RR6BL0000204	F	69	II	4	1	1568	Yes	7.018526316	3.94	5.853328	3.05
RR6BL0000205	F	72	II	16	0	1568	Yes	5.709106061	2.17	4.937048	2.41
RR6BL0000213	F	58	II	4	0	1542	Yes	5.550716667	3.55	5.100257	3.01

Appendix 1b

RR6BL0000214	F	68	III	6	0	1052	No	8.432573529	3.29	6.325548	3.34
RR6BL0000225	F	67	I	4	0	1140	Yes	4.143132813	2.24	4.175458	1.55
RR6BL0000230	F	43	II	9	2	1533	Yes	5.615145695	3.11	5.785374	2.44
RR6BL0000232	F	36	II	15	2	1369	Yes	4.867612903	2.91	4.816052	2.54
RR6BL0000235	F	53	III	14	3	1477	Yes	3.93407483	2.78	3.969718	2.24
RR6BL0000245	F	86	II	3	1	1244	Yes	3.950277778	2.19	4.453422	2.59
RR6BL0000252	F	52	III	19	19	889	Yes	4.673604444	3.25	5.270442	2.81
RR6BL0000259	F	75	III	8	4	633	No	3.69161809	2.73	3.805464	2.60
RR6BL0000274	F	79	I	1	0	1397	Yes	3.339478528	2.59	4.804747	2.08
RR6BL0000350	F	59	II	3	0	1175	Yes	4.053057554	2.63	4.161381	1.70
RR6BL0000386	F	75	II	2	0	1129	Yes	4.064532663	2.42	4.227759	2.11
RR6BL0000409	F	43	III	20	2	1281	Yes	4.866966942	2.98	4.939158	2.68
RR6BL0000528	F	56	II	3	0	670	Yes	7.240571429	3.41	6.219889	4.47
RR6BL0000535	F	46	III	3	0	710	Yes	3.693757143	2.53	3.837276	1.64
RR6BL0000540	F	44	III	3	1	696	Yes	5.318149606	3.12	4.11476	3.29
RR6BL0000557	F	83	III	3	0	723	Yes	4.868805195	2.99	5.138013	2.88
RR6BL0000571	F	42	III	15	1	584	Yes	4.432066667	2.60	4.490789	2.38
RR6BL0000623	F	48	III	2	0	339	Yes	4.263055901	2.35	4.856321	3.46
RVCC40000001	F	71	III	16	0	2313	Yes	3.693186275	1.89	4.904157	1.90
RVCC40000011	F	58	II	5	1	2282	No	1.075	2.23	2.895674	1.38
RVCC40000015	F	87	III	17	0	2157	Yes	3.527236111	1.55	4.189458	2.51
RVCC40000016	F	60	II	30	3	2268	Yes	7.00902963	3.54	6.317881	4.30
RVCC40000020	F	55	III	19	0	2058	Yes	2.484841667	1.92	2.923891	1.42
RVCC40000021	F	63	II	20	0	2006	Yes	7.824867925	4.85	6.80783	6.16
RVCC40000022	F	79	III	27	1	143	No	2.182	3.44	5.185153	2.20
RVCC40000026	F	45	I	5	0	2233	Yes	4.844267717	3.09	3.779026	1.65
RVCC40000029	F	58	III	26	3	2228	No	3.493176471	1.49	4.155758	2.95
RVCC40000031	F	54	III	32	0	2232	Yes	4.172952381	2.15	3.208333	1.65
RVCC40000035	F	68	III	16	3	2137	Yes	5.175631579	2.11	3.215313	1.08

Appendix 1b

RVCC40000039	F	46	III	15	0	2221	No	2.659	0.99	2.487071	1.51
RVCC40000063	F	81	III	26	2	1897	Yes	2.766083333	2.09	3.349462	1.95
RVCC40000067	F	48	III	15	1	2286	No				
RVCC40000070	F	73	II	35	0	2077		3.806033898	2.30	3.953957	1.77
RVCC40000071	F	85	III	17	3	2060	Yes	2.300756522	1.72	5.102733	1.10
RVCC40000072	F	65	III	25	0	1973	Yes	3.781486486	3.28	3.753628	2.43
RVCC40000076	F	43	III	28	6	1171	Yes	1.845	2.63	4.614727	2.40
RVCC40000089	F	66	III	14	2	2022	Yes	5.820268293	3.93	5.040171	2.61
RVCC40000090	F	84	II	18	0	2113	Yes	4.440637255	2.68	5.4594	2.68
RVCC40000124	F	45	II	19	3		Yes	2.919257576	1.20	4.360538	2.56
RVCC40000129	F	56	II	4	0	2000	Yes	4.031826087	2.52	3.564782	1.88
RVCC40000134	F	70	III	17	0	1948	Yes	2.430598291	1.16	2.49013	0.98
RVCC40000135	F	83	I	19	0	2142	Yes	4.663242105	2.93	5.01588	2.41
RVCC40000154	F	49	III	21	1	1921	Yes	3.237545455	1.51	2.78837	1.33
RVCC40000201	F	66	III	17	1	1885	No	4.902689873	2.80	4.249661	2.14
RVCC40000348	F	87	I	21	0	1866	Yes	4.082888889	2.61	4.187718	2.30
RVCC40000555	F	76	III	21	20	1151	No	3.325854167	1.30	3.649698	0.97
RVCC40000559	F	61	III	4	1	1638	No	5.111390374	3.42	4.24685	2.59
RVCC40000591	F	50	III	5	1	484	Yes	3.197876543	1.22	3.689857	0.86
RVCC40000597	F	48	II	20	3	1589	Yes	2.047924731	0.86	3.485328	1.32
RVCC40000622	F	59	III	3	0	1477	Yes	2.674352381	1.49	2.519857	1.48
RVCC40000631	F	79	III	23	0	1492	Yes	1.776765625	1.28	4.900139	1.52
RVCC40000633	F	68	II	6	0	1563	Yes	2.918557692	1.13	3.571769	1.05
RVCC40000638	F	91	II	6	1		No	3.034938931	1.61	4.872512	1.33
RVCC40000642	F	70	II	6	0	1479	No	3.126026316	1.02	3.852204	1.69
RVCC40000643	F	42	II	5	0	1561	Yes	3.13050838	1.81	4.434177	2.63
RVCC40000665	F	65	II	5	2	1570	Yes	4.135612903	2.49	3.86658	2.22
RVCC40000686	F	49	III	5	1	1424	Yes	3.407466667	1.27	2.845194	1.51
RVCC40000687	F	49	II	6	0	1527	Yes	2.234923077	1.02	5.413813	2.92

Appendix 1b

RVCC40000696	F	42	III	26	14		Yes	1.791929825	1.22	3.522867	2.05
RVCC40000737	F	78	III	12	0	1354	Yes	2.642169231	1.41	4.034481	1.07
RVCC40000752	F	47	II	1	0	1442	Yes	3.865352941	1.58	3.137836	1.25
RVCC40000753	F	68	III	2	0	1423	Yes	1.664		1.896	
RVCC40000776	F	65	II	4	0	1439	Yes	4.19183871	2.35	3.086189	2.40
RVCC40000786	F	60	II	4	0	1274	Yes	4.118711111	2.17	4.482837	1.82
RVCC40000795	F	65	I	2	2	1408	Yes	4.083505155	1.37	4.753698	5.62
RVCC40000907	F	53	II	4	0	1128	Yes	3.802625	2.55	4.651324	2.18
RVCC40000936	F	61	II	2	1	1231	Yes	5.784782353	2.99	5.909642	3.43
RVCC40000965	F	73	III	3	0	1265	Yes	1.636529412	0.62	1.936333	1.41
RVCC40001036	F	73	III	30	3	985	Yes	2.58975	0.85	1.654438	1.60
RVCC40001055	F	50	II	2	0	1295	Yes	2.633237288	1.47	2.461315	1.46
RVCC40001060	F	45	III	25	5	991	Yes	3.736214876	3.35	4.516152	2.14
RVCC40001111	F	48	III	4	0	1142	Yes	5.451528986	2.86	6.883327	3.68
RVCC40001236	F	55	III	2	0	896	Yes	4.092526316	2.34	3.819763	2.73
RVCC40001469	F	81	I	3	0	646	Yes	3.851755102	1.36	1.819808	1.49

Appendix 1b

wcbnum	NPI Score	Breast Grade	ER Score	pgr score	her2score	Her2 2+ : Pos or neg following FISH	Alive	Died of Cancer	OS	XpYp tel
RR6BL0000011	0	III	Negative				No	Yes	233	1.075
RVCC40000033	6.82	III	Negative	Negative	3	positive	No	Yes	852	1.290
RVCC40000753	4.36	III	positive		0	negative	Yes		1423	1.664
RR6BL0000004	0	II	positive				No	Yes	241	1.780
RR6BL0000120	0	II					No	Yes	396	1.818
RVCC40000076	6.46	III	Negative	Negative	0	negative	No	Yes	1760	1.845
RR6BL0000001	0	III	positive	Negative	0	negative	No	Yes	300	2.021
RVCC40000022	5.42	III	Negative	Negative	3	positive	No	Yes	241	2.182
RR6BL0000008	0	II	positive				Yes		2467	2.290
RVCC40001036	5.84	III	Negative	Negative	2	negative	Yes		985	2.328
RR6BL0000350	0	II	positive	positive	1	negative	Yes		1564	2.458
RR6BL0000083	0	III	positive				No	Yes	2015	2.512
RR6BL0000274	0	I	positive		0	negative	Yes		1766	2.517
RVCC40000124	4.8	II	positive	positive	0	negative	Yes		1997	2.617
RVCC40000622	4.44	III	positive	positive	0	negative	Yes		1477	2.660
RVCC40000965	4.72	III	Negative	Negative	3	positive	Yes		1265	2.669
RR6BL0000013	0	III	Negative	Negative	2	negative	Yes		2305	2.680
RVCC40000021	3.28	II	positive		0	negative	Yes		2006	2.686
RR6BL0000039	0	III	Negative	positive			Yes		2388	2.700
RVCC40000665	0	II	positive		0	negative	Yes		1570	2.701
RVCC40000072	4.36	III	Negative	positive	0	negative	Yes		1973	2.729
RR6BL0000045	0	II	positive				Yes		2407	2.730
RVCC40000907	4.54	II	positive		0	negative	Yes		1128	2.793
RR6BL0000067	0	III	positive	Negative	2	negative	Yes		2271	2.800
RR6BL0000072	0	III	Negative	Negative	3	positive	Yes		2264	2.920
RVCC40000135	2.4	I	positive				Yes		2142	2.946
RVCC40000009	6.6	III	positive		3	positive	No	Unknown		2.960
RVCC40000025	4.3	III	positive		3	positive	Yes		2179	2.980
RVCC40000030	3.34	II	positive				Yes		2109	3.040
RR6BL0000386	0	II	positive	positive	2	negative	Yes		1493	3.044
RR6BL0000002	0	II	Negative	Negative			Yes		2306	3.118
RVCC40000752	3.4	II	positive		1	negative	Yes		1442	3.153

RR6BL0000020	0	II	positive				Yes		2432	3.158
RVCC40001469	2.6	I	positive	positive	1	negative	Yes		646	3.225
RVCC40000042	3.5	II	positive				No	Unknown		3.230
RVCC40000144	4.6	III	Negative	Negative	1	negative	Yes		2148	3.260
RVCC40000051	2.24	I	positive				Yes		1977	3.260
RR6BL0000171	0	II	positive		1	negative	Yes		2060	3.265
RVCC40001055	3.28	II	positive		1	negative	Yes		1295	3.353
RVCC40000696	6.46	III	Negative	Negative	3	positive	Yes		1329	3.366
RVCC40000089	5.26	III	positive	positive	0	negative	Yes		2022	3.434
RVCC40000155	5.48	III	positive				Yes		1927	3.450
RVCC40000165	5.6	III	positive		0	negative	Yes		1917	3.500
RVCC40000795	3.56	I	positive		1	negative	Yes		1408	3.520
RVCC40000070	3.56	II	Negative	Negative	1	negative	Yes		2077	3.542
RVCC40000029	5.02	II	positive	positive	0	negative	Yes		2228	3.554
RR6BL0000104	0	III	positive	Negative	2	negative	Yes		2141	3.580
RVCC40000631	4.66	III	positive		3	positive	Yes		1492	3.602
RVCC40000166	4.34	III	Negative		3	positive	Yes		1922	3.700
RVCC40000936	5.08	II	positive		0	negative	Yes		1231	3.702
RR6BL0000111	0	II	positive	Negative			No	No		3.719
RVCC40000035	5.46	III	positive	positive	0	negative	Yes		2137	3.720
RR6BL0000232	0	II	positive	Negative	3	positive	Yes		1749	3.721
RVCC40000016	4.54	II	positive	positive	2	negative	Yes		2268	3.748
RR6BL0000204	0	II	positive	positive	0	negative	Yes		1935	3.763
RVCC40000026	2.28	I	positive	positive	0	negative	Yes		2233	3.778
RVCC40000633	3.44	II	positive	positive	2	negative	Yes		1563	3.808
RR6BL0000252	0	III	positive	positive	1	negative	No	Yes	1754	3.822
RVCC40000686	5.52	III	positive		0	negative	Yes		1414	3.830
RVCC40000039	4.6	III	Negative	Negative	0	negative	Yes		2221	3.833
RVCC40000063	6.3	III	Negative	Negative	1	negative	No	Unknown		3.879
RVCC40000642	3.28	II	positive	positive	0	negative	Yes		1479	3.892
RR6BL0000557	0	III	positive	positive	3	positive	Yes		723	3.961
RR6BL0000070	0	III	Negative	Negative	3	positive	No	Unknown		3.977
RVCC40001236	4.36	III	positive	positive	0	negative	Yes		896	3.981
RR6BL0000623	0	III	positive	Negative	3	positive	Yes		721	3.987
RR6BL0000214	0	III	positive	positive	0	negative	No	Yes	766	4.018

RVCC40000015	4.66	III	positive	Negative	3	positive	Yes		2157	4.059
RVCC40000539	3.3	II	positive	positive	0	negative	Yes		1362	4.080
RVCC40000134	4.42	III	positive	Negative	0	negative	Yes		1948	4.080
RR6BL0000005	0	II	positive		1	negative	Yes		1725	4.092
RVCC40000573	3.96	I	positive	positive	0	negative	Yes		1436	4.100
RVCC40000011	4.44	II	positive		2	negative	Yes		2282	4.110
RVCC40000001	4.44	III	Negative	Negative	3	positive	Yes		2313	4.111
RR6BL0000198	0	III	positive		3	positive	No	Yes	195	4.120
RVCC40001111	4.44	III	positive	positive	1	negative	Yes		1142	4.124
RR6BL0000186	0	I	positive	positive	1	negative	No	Yes	1792	4.146
RR6BL0000019	0	I	positive		1	negative	Yes		2436	4.163
RR6BL0000528	0	II	positive	Negative	3	positive	Yes		1048	4.170
RVCC40000638	0	II	positive	positive	0	negative	No	Unknown		4.200
RVCC40000583	5.98	II	positive	positive	2	negative	Yes		1508	4.200
RVCC40000559	4.24	II	positive	positive	2	negative	Yes		1638	4.219
RVCC40000589	6.8	III	positive	positive	0	negative	Yes		1502	4.220
RVCC40000632	5.76	III	Negative	Negative	0	negative	Yes		1523	4.230
RR6BL0000117	0	II	positive		1	negative	Yes		2173	4.290
RR6BL0000141	0	III	Negative		3	positive	No	Yes	854	4.378
RR6BL0000213	0	II	positive		3	positive	Yes		1907	4.459
RR6BL0000033	0	II	positive				No	No		4.468
RVCC40000647	4.44	III	Negative	Negative	2	negative	Yes		1442	4.500
RR6BL0000245	0	II	positive	positive	0	negative	Yes		1625	4.545
RR6BL0000064	0	II	positive	positive	1	negative	Yes		2271	4.635
RVCC40000031	4.32	III	Negative	Negative	1	negative	Yes		2232	4.659
RR6BL0000184	0	III	Negative	Negative	0	negative	No	Yes	1000	4.659
RVCC40000687	3.44	II	positive		1	negative	Yes		1527	4.690
RVCC40000649	5.62	III	Negative	Negative	1	negative	Yes		1440	4.700
RVCC40000597	4.6	II	positive	positive	0	negative	Yes		1589	4.701
RVCC40000067	5.34	III	positive	positive	1	negative	Yes		2286	4.866
RVCC40000591	5.66	III	positive	positive	1	negative	No	Yes	546	4.871
RR6BL0000199	0	II	positive	positive	1	negative	Yes		2095	4.894
RR6BL0000535	0	III	positive	positive	3	positive	Yes		1085	4.895
RR6BL0000235	0	III	positive	Negative	3	positive	Yes		1858	4.896
RR6BL0000205	0	II	positive	positive	0	negative	Yes		1935	4.930

RVCC40000786	3.44	II	positive				Yes		1274	4.991
RVCC40000776	3.24	II	positive		0	negative	Yes		1439	5.138
RR6BL0000131	0	II	positive	positive	1	negative	Yes		2131	5.146
RR6BL0000409	0	III	positive	positive	3	positive	Yes		1281	5.151
RVCC40000348	2.44	I	positive	positive	0	negative	Yes		1866	5.203
RVCC40000698	4.82	III	Negative	Negative	0	negative	No	Yes	412	5.300
RR6BL0000230	0	II	positive	positive	1	negative	Yes		1913	5.345
RVCC40000555	6.64	III	positive	Negative	0	negative	No	Yes	1233	5.421
RVCC40000643	3.26	II	positive		2	positive	Yes		1561	5.463
RR6BL0000202	0	II	positive		0	negative	Yes		1884	5.479
RVCC40000090	3.4	II	positive				Yes		2113	5.519
RR6BL0000571	0	III	positive	positive	3	positive	Yes		959	5.578
RVCC40000755	4.28	III	positive		2	negative	Yes		1307	5.600
RR6BL0000540	0	III	positive	Negative	3	positive	Yes		1071	5.642
RVCC40000201	5.7	III	Borderline	Borderline	3	positive	Yes		1885	5.736
RVCC40000154	6	III	Negative	Negative	2	negative	Yes		1921	5.750
RVCC40000758	6.58	III	positive		0	negative	Yes		1328	5.800
RR6BL0000134	0	II	positive	positive	0	negative	Yes		2122	5.812
RR6BL0000071	0	II	Negative	Negative	0	negative	No	Yes	719	5.847
RR6BL0000259	0	III	positive	Negative	3	positive	No	Yes	678	6.130
RVCC40000737	4.5	III	positive	Borderline	2	negative	Yes		1344	7.036
RVCC40001060	0	III	Negative	Negative	3	positive	Yes		991	7.046
RR6BL0000225	0	I	positive	positive	0	negative	Yes		1507	7.268
RVCC40000071	6.6	III	positive	Negative			No	Unknown		7.852
RR6BL0000130	0	III	Negative	Negative	2	negative	Yes		2145	8.460
RVCC40000020	4.44	III	Negative	Negative	3	positive	Yes		2058	9.975
RR6BL0000170	0	I	positive	positive	1	negative	Yes		2060	
RVCC40000129	3.6	II	positive	positive	0	negative	Yes		2000	

Table A Summary of Grade and node status along side mean telomere length and Standard deviation at XpYp and 17p for all IDC patients

wcbnum	Age at Operation	pTNM	NPI Score	Breast Grade	ER Score	pgr score	her2score	Her2 2+ : Pos or neg following FISH	XpYp tel
RR6BL0000011	84		0	III	Negative				1.075

Appendix 1b

RVCC40000033	75	T2 N3a Mx	6.82	III	Negative	Negative	3	positive	1.290
RVCC40000753	68	T1c No Mx	4.36	III	positive		0	negative	1.664
RR6BL0000004	82	T2 No Mx	0	II	positive				1.780
RR6BL0000120	69		0	II					1.818
RVCC40000076	43	T2 N2a Mx	6.46	III	Negative	Negative	0	negative	1.845
RR6BL0000001	72	T2 N1 Mx	0	III	positive	Negative	0	negative	2.021
RVCC40000022	79	T2N1miMx	5.42	III	Negative	Negative	3	positive	2.182
RR6BL0000008	61	T1 No Mx	0	II	positive				2.290
RVCC40001036	73	T2N1Mx	5.84	III	Negative	Negative	2	negative	2.328
RR6BL0000350	59		0	II	positive	positive	1	negative	2.458
RR6BL0000083	75	T2 N1 Mx	0	III	positive				2.512
RR6BL0000274	79	T1 No	0	I	positive		0	negative	2.517
RVCC40000124	45	T2N1aMx	4.8	II	positive	positive	0	negative	2.617
RVCC40000622	59	T2N0Mx	4.44	III	positive	positive	0	negative	2.660
RVCC40000965	73	T2NOMX	4.72	III	Negative	Negative	3	positive	2.669
RR6BL0000013	48	T1 No Mx	0	III	Negative	Negative	2	negative	2.680
RVCC40000021	63	T1c No Mx	3.28	II	positive		0	negative	2.686
RR6BL0000039	34	T2 No Mx	0	III	Negative	positive			2.700
RVCC40000665	65		0	II	positive		0	negative	2.701
RVCC40000072	65	T1c No Mx	4.36	III	Negative	positive	0	negative	2.729
RR6BL0000045	56	T2 No Mx	0	II	positive				2.730
RVCC40000907	53	T2mN1Mx	4.54	II	positive		0	negative	2.793
RR6BL0000067	58	T2 No Mx	0	III	positive	Negative	2	negative	2.800
RR6BL0000072	58	T2 No Mx	0	III	Negative	Negative	3	positive	2.920
RVCC40000135	83	T1cNoMx	2.4	I	positive				2.946
RVCC40000009	68	T2 N2a Mx	6.6	III	positive		3	positive	2.960
RVCC40000025	38		4.3	III	positive		3	positive	2.980
RVCC40000030	74	T1cNoMx	3.34	II	positive				3.040
RR6BL0000386	75	t1n0mx	0	II	positive	positive	2	negative	3.044

Appendix 1b

RR6BL0000002	73	T2 N1 Mx	0	II	Negative	Negative			3.118
RVCC40000752	47	T2N0Mx	3.4	II	positive		1	negative	3.153
RR6BL0000020	86		0	II	positive				3.158
RVCC40001469	81	T2NoMx	2.6	I	positive	positive	1	negative	3.225
RVCC40000042	85	T2 No Mx	3.5	II	positive				3.230
RVCC40000144	39	T2N0Mx	4.6	III	Negative	Negative	1	negative	3.260
RVCC40000051	42	T1c No Mx	2.24	I	positive				3.260
RR6BL0000171	45	T2 N1	0	II	positive		1	negative	3.265
RVCC40001055	50	T1cNoMx	3.28	II	positive		1	negative	3.353
RVCC40000696	42	T2N3aMx	6.46	III	Negative	Negative	3	positive	3.366
RVCC40000089	66	T1c N1 Mx	5.26	III	positive	positive	0	negative	3.434
RVCC40000155	78	T2 N1 Mx	5.48	III	positive				3.450
RVCC40000165	47	T2 N1m Mx	5.6	III	positive		0	negative	3.500
RVCC40000795	65	T2N1aMx	3.56	I	positive		1	negative	3.520
RVCC40000070	73	T2 No Mx	3.56	II	Negative	Negative	1	negative	3.542
RVCC40000029	58	T3 N1 Mx	5.02	II	positive	positive	0	negative	3.554
RR6BL0000104	59	T2 No Mx	0	III	positive	Negative	2	negative	3.580
RVCC40000631	79	T2N0Mx	4.66	III	positive		3	positive	3.602
RVCC40000166	69	T1c No Mx	4.34	III	Negative		3	positive	3.700
RVCC40000936	61	T3N1miMx	5.08	II	positive		0	negative	3.702
RR6BL0000111	79	T1 N1 Mx	0	II	positive	Negative			3.719
RVCC40000035	68	T2 N1a Mx	5.46	III	positive	positive	0	negative	3.720
RR6BL0000232	36		0	II	positive	Negative	3	positive	3.721
RVCC40000016	60	T2 N12 Mx	4.54	II	positive	positive	2	negative	3.748
RR6BL0000204	69		0	II	positive	positive	0	negative	3.763
RVCC40000026	45	T1c No Mo	2.28	I	positive	positive	0	negative	3.778
RVCC40000633	68	T2aNoMx	3.44	II	positive	positive	2	negative	3.808
RR6BL0000252	52		0	III	positive	positive	1	negative	3.822
RVCC40000686	49	T2N1miMx	5.52	III	positive		0	negative	3.830

RVCC40000039	46	pT2 No Mx	4.6	III	Negative	Negative	0	negative	3.833
RVCC40000063	81	T3 N1 Mx	6.3	III	Negative	Negative	1	negative	3.879
RVCC40000642	70	T4bNoMx	3.28	II	positive	positive	0	negative	3.892
RR6BL0000557	83	T2N0MX	0	III	positive	positive	3	positive	3.961
RR6BL0000070	80	T2 N2 Mx	0	III	Negative	Negative	3	positive	3.977
RVCC40001236	55	T1cN0Mx	4.36	III	positive	positive	0	negative	3.981
RR6BL0000623	48	t2n0mx	0	III	positive	Negative	3	positive	3.987
RR6BL0000214	68		0	III	positive	positive	0	negative	4.018
RVCC40000015	87	T2NoMx	4.66	III	positive	Negative	3	positive	4.059
RVCC40000539	70	T1cN0Mx	3.3	II	positive	positive	0	negative	4.080
RVCC40000134	70	T2 No Mx	4.42	III	positive	Negative	0	negative	4.080
RR6BL0000005	53	T2 N1 Mx	0	II	positive		1	negative	4.092
RVCC40000573	46	T2NoMx	3.96	I	positive	positive	0	negative	4.100
RVCC40000011	58	T2 N1a Mx	4.44	II	positive		2	negative	4.110
RVCC40000001	71	T2 No Mx	4.44	III	Negative	Negative	3	positive	4.111
RR6BL0000198	51	T2 N2 Mx	0	III	positive		3	positive	4.120
RVCC40001111	48	T2N0Mx	4.44	III	positive	positive	1	negative	4.124
RR6BL0000186	80		0	I	positive	positive	1	negative	4.146
RR6BL0000019	49	T2 No Mx	0	I	positive		1	negative	4.163
RR6BL0000528	56	T1N0MX	0	II	positive	Negative	3	positive	4.170
RVCC40000638	91	T2N1	0	II	positive	positive	0	negative	4.200
RVCC40000583	47	T2N2aMx	5.98	II	positive	positive	2	negative	4.200
RVCC40000559	61	T1cN1aMx	4.24	II	positive	positive	2	negative	4.219
RVCC40000589	76	T2mN3aMx	6.8	III	positive	positive	0	negative	4.220
RVCC40000632	33	T2N1Mx	5.76	III	Negative	Negative	0	negative	4.230
RR6BL0000117	67	T2 N1 Mx	0	II	positive		1	negative	4.290
RR6BL0000141	48		0	III	Negative		3	positive	4.378
RR6BL0000213	58	T1 No	0	II	positive		3	positive	4.459
RR6BL0000033	84	T2 N1 Mx	0	II	positive				4.468

Appendix 1b

RVCC40000647	65	T2N0Mx	4.44	III	Negative	Negative	2	negative	4.500
RR6BL0000245	86		0	II	positive	positive	0	negative	4.545
RR6BL0000064	40	T1 No Mx	0	II	positive	positive	1	negative	4.635
RVCC40000031	54	T1c No Mx	4.32	III	Negative	Negative	1	negative	4.659
RR6BL0000184	42	T3N0MX	0	III	Negative	Negative	0	negative	4.659
RVCC40000687	49	T2 No Mx	3.44	II	positive		1	negative	4.690
RVCC40000649	60	T2N1aMx	5.62	III	Negative	Negative	1	negative	4.700
RVCC40000597	48	T2N1aMx	4.6	II	positive	positive	0	negative	4.701
RVCC40000067	48	T1 N1 Mo	5.34	III	positive	positive	1	negative	4.866
RVCC40000591	50	T2N1Mx	5.66	III	positive	positive	1	negative	4.871
RR6BL0000199	70	T1 No Mx	0	II	positive	positive	1	negative	4.894
RR6BL0000535	46	T2N0MX	0	III	positive	positive	3	positive	4.895
RR6BL0000235	53	T2	0	III	positive	Negative	3	positive	4.896
RR6BL0000205	72	Unknown	0	II	positive	positive	0	negative	4.930
RVCC40000786	60	T2NoMx	3.44	II	positive				4.991
RVCC40000776	65	T1cN0Mx	3.24	II	positive		0	negative	5.138
RR6BL0000131	69		0	II	positive	positive	1	negative	5.146
RR6BL0000409	43	t2n1m0	0	III	positive	positive	3	positive	5.151
RVCC40000348	87	T2NoMx	2.44	I	positive	positive	0	negative	5.203
RVCC40000698	70	T2N0Mx	4.82	III	Negative	Negative	0	negative	5.300
RR6BL0000230	43		0	II	positive	positive	1	negative	5.345
RVCC40000555	76	T2 N3a Mx	6.64	III	positive	Negative	0	negative	5.421
RVCC40000643	42	T1N0Mx	3.26	II	positive		2	positive	5.463
RR6BL0000202	66		0	II	positive		0	negative	5.479
RVCC40000090	84	T1c No Mx	3.4	II	positive				5.519
RR6BL0000571	42	T2N1biMO	0	III	positive	positive	3	positive	5.578
RVCC40000755	73	T2N0Mx	4.28	III	positive		2	negative	5.600
RR6BL0000540	44	T1cN1biMX	0	III	positive	Negative	3	positive	5.642
RVCC40000201	66	T2 N1 Mx	5.7	III	Borderline	Borderline	3	positive	5.736

Appendix 1b

RVCC40000154	49	T2 N1 Mx	6	III	Negative	Negative	2	negative	5.750
RVCC40000758	52	T2N3aMx	6.58	III	positive		0	negative	5.800
RR6BL0000134	68		0	II	positive	positive	0	negative	5.812
RR6BL0000071	71	T2 N1 Mx	0	II	Negative	Negative	0	negative	5.847
RR6BL0000259	75		0	III	positive	Negative	3	positive	6.130
RVCC40000737	78	T4bNoMx	4.5	III	positive	Borderline	2	negative	7.036
RVCC40001060	45	T3N2Mx	0	III	Negative	Negative	3	positive	7.046
RR6BL0000225	67	Unknown	0	I	positive	positive	0	negative	7.268
RVCC40000071	85	T3 N1	6.6	III	positive	Negative			7.852
RR6BL0000130	47	T1 No Mx	0	III	Negative	Negative	2	negative	8.460
RVCC40000020	55	T2 No Mx	4.44	III	Negative	Negative	3	positive	9.975
TOTAL +ive				ER	94		HER	29	

Table B summary of TNM stage, HER2 status, ER/PR status and grade for each patient alongside mean telomere length at XpYp and 17p

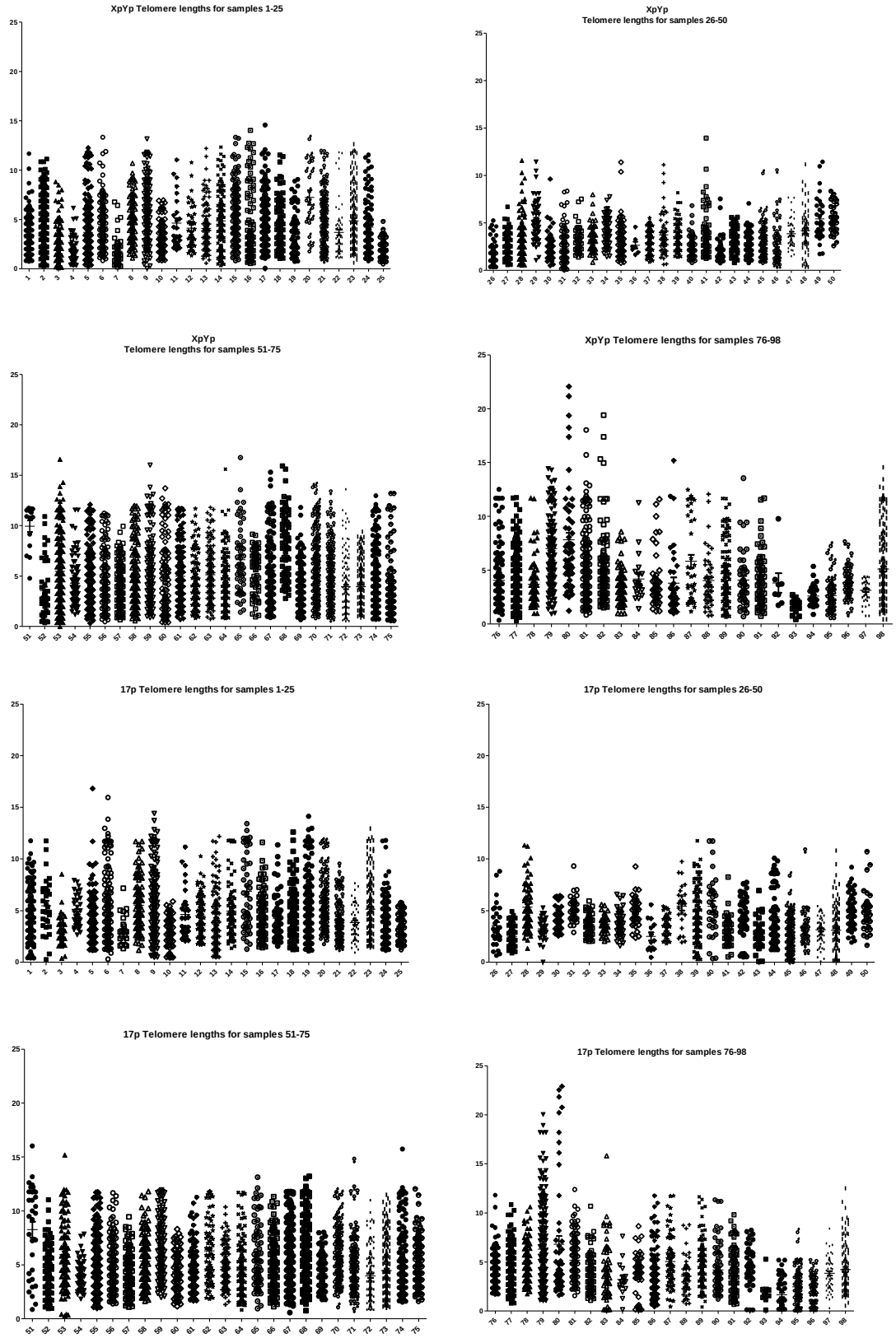
Appendix 2 Table B
Telomere standard deviations with HER2 status

wcbnum	Her2 2+ : Pos or neg following FISH	XpYp tel		SD
RR6BL0000070	positive	3.977		**4.01**
RR6BL0000072	positive	2.920		**3.25**
RR6BL0000141	positive	4.378		**3.10**
RR6BL0000198	positive	4.120		**2.62**
RR6BL0000213	positive	4.459		**3.29**
RR6BL0000232	positive	3.721		**2.78**
RR6BL0000235	positive	4.896		**2.19**
RR6BL0000259	positive	6.130		**2.59**
RR6BL0000409	positive	5.151		**3.41**
RR6BL0000528	positive	4.170		**2.53**
RR6BL0000535	positive	4.895		**3.12**
RR6BL0000540	positive	5.642		**2.99**
RR6BL0000557	positive	3.961		**2.60**
RR6BL0000571	positive	5.578		2.35
RR6BL0000623	positive	3.987		1.89
RVCC40000001	positive	4.111		1.78
RVCC40000009	positive	2.960		2.23
RVCC40000015	positive	4.059		3.54
RVCC40000020	positive	9.975		4.85
RVCC40000022	positive	2.182		1.25
RVCC40000025	positive	2.980		3.09
RVCC40000033	positive	1.290		2.11
RVCC40000166	positive	3.700		2.80
RVCC40000201	positive	5.736		2.61
RVCC40000631	positive	3.602		1.11
RVCC40000643	positive	5.463		1.25
RVCC40000696	positive	3.366		1.82
RVCC40000965	positive	2.669		0.85
RVCC40001060	positive	7.046		2.86

Table B: shows all patients that tested positive for HER2 status (using *FISH)*. Shown in bold are all those patients of this subgrouping who had standard deviation telomere lengths of 2.5kb and above

Appendix 3: Scatter plot A of all STELA telomere lengths

Appendix 1b



Scatter plot A: shows all telomere lengths represented using scatter diagrams and subdivided by telomere (XpYp and 17p)

Appendix 4 Table C: Fusion frequencies
Diploid genome equivalents = 5001 (30ng) per reaction (166.7 molecules/ng)
18 reactions = 90000 input genome equivalents

No.	1	2	3	4	5	6	7	8
17p	0	2.2×10^{-5}	1.1×10^{-5}	0	1.1×10^{-5}	3.3×10^{-5}	0	1.1×10^{-5}
21q	0	0	0	1.1×10^{-5}	0	0	1.1×10^{-5}	0
16p	1.1×10^{-5}	0	1.1×10^{-5}	2.2×10^{-5}	0	0	2.2×10^{-5}	0
No.	8	9	10					
17p	0	1.1×10^{-5}	0					
21q	1.1×10^{-5}	0	2.2×10^{-5}					
16p								
No.	11	12	13	14	15	16	17	
XpYp	0	1.1×10^{-5}	0	0	1.1×10^{-5}	0	0	
17p	0	0	2.2×10^{-5}	0	0	1.1×10^{-5}	0	
21q	1.1×10^{-5}	2.2×10^{-5}	2.2×10^{-5}		0	1.1×10^{-5}	2.2×10^{-5}	
16p	0	1.1×10^{-5}	1.1×10^{-5}	0	0	0	0	
No.	18	19	20	21	22			
17p	0	1.1×10^{-5}	0	0	0			
21q	0	0	2.2×10^{-5}	1.1×10^{-5}	0			
16p	2.2×10^{-5}	0	0	0	1.1×10^{-5}			
No.	23	24						
17p	2.1×10^{-5}	1.1×10^{-5}						

Table C: shows the fusion frequencies calculated on a patient by patient basis. ‘No.’ refers to the patient number. Only patients with fusion banding patterns have been included here, and they are subdivided by telomere that was probed for. XpYp was detected using the XpYpo-g probe; 17p with 17p6; 21q with 21qseq1rev and 16p with 16p1.