# DNA MISMATCH REPAIR DEFICIENCY AND TELOMERE DYNAMICS 

$\mathrm{M}^{\mathrm{a}}$ Carmen Garrido Navas
Department of Genetics

This dissertation is submitted for the degree of
Doctor of Philosophy


#### Abstract

DNA mismatch repair (MMR) deficiency is the hallmark of Lynch syndrome (LS), an autosomal inherited predisposition for an early onset of colorectal cancer (CRC) characterized by the presence of high levels of microsatellite instability. CRC cell lines carrying a MMR gene mutation show higher mutation rates than MMR-proficient cells not only at microsatellites but also at telomeres.

Here, the MMR protein MSH2 was downregulated in WI38 cells using an shRNA and stable clones were grown in a $5 \% \mathrm{O}_{2}$ environment to reduce the effect of oxidative stress on the telomere shortening rate (TSR). It was found that clones with a greater than $60 \%$ protein downregulation had a higher TSR with an exponential relationship between protein content and TSR for the B but not the $\Delta$ allele for 12 q what might be related with the telomeric variant repeat structure. Additionally, oxidative stress was found to have a synergistic effect on TSR together with MMR deficiency. Mutant XpYp telomeres generated from LoVo ( $\mathrm{MSH}^{-/-}$) were studied by telomere variant repeat (TVR) analysis to understand the mutation process and a tendency towards deletions was suggested. Furthermore, a novel $3^{\text {rd }}$ next generation sequencing technology was tested to obtain full information of the telomere repeat array establishing the bases for future experiments. Finally, telomere length was measured in saliva DNA from 37 controls and 91 LS patients. It was found that telomeres shortened with age at similar rates in both cohorts and no significant relationship was found between telomere length and age for $\mathrm{MLH} 1^{+/-}$ patients. Besides, $\mathrm{MSH}^{+/-}$patients had significantly shorter age-adjusted telomere length for XpYp but not 12 q than $\mathrm{MMR}^{+/+}$and parent-children pair comparisons for age-adjusted telomere lengths showed that $\mathrm{MMR}^{+/+}$children had longer telomeres than their $\mathrm{MMR}^{+/-}$siblings and their $\mathrm{MMR}^{+/-}$parent but shorter than their $\mathrm{MMR}^{+/+}$parent.


## Acknowledgements

I would firstly like to thank my supervisor Dr. Nicola Royle and the Department of Genetics (Univ. of Leicester) for giving me the opportunity of working in this project. Also, to my Thesis committee members Dr. Salvador Macip and Dr. Steven Foster for their support and useful feedback.

I cannot forget and I want to thank all the people who helped me at some point with the labwork: Dr. Yan Huang with STELA and TRAP, Dr. Victoria Cotton with TVR, Dr. Rachel Turner with WB, Dr. Richard Badge for suggestions on the sequencing analysis and Rita Neumann with the minION ${ }^{\text {TM }}$ sequencing and dot blots. I also want to thank all the technician team, specially Nicky and Milly, for their constant help. I will also remember the students that I have worked with and learned from: Despina, Dr. Frances, Emma and Matt. I am also very grateful for the friends that made my life much more enjoyable in the lab: Maria Pakendorf, Enjie Zhang, Dr. Nick Eastley and Ihthi Ali.

Finally, a special praise for my parents and my partner without whose unconditional support I wouldn't have been able to achieve this project, THANKS for your patience and love.

## Contents

ABSTRACT ..... i
ACKNOWLEDGEMENTS ..... ii
ABBREVIATIONS ..... xi
1 Introduction ..... 1
1.1 DNA mismatch repair ..... 1
1.1.1 Origin of mismatches in the DNA ..... 1
1.1.2 MMR mechanism in prokaryotes ..... 3
1.1.2.1 MMR in E. coli ..... 3
1.1.3 MMR mechanism in eukaryotes ..... 8
1.1.3.1 MMR mechanism in humans ..... 8
1.1.3.2 Models for MMR mechanism ..... 14
1.1.4 MMR, MSI and cancer predisposition ..... 17
1.1.5 Other functions for MMR ..... 19
1.2 Telomeres ..... 23
1.2.1 Function and structure of human telomeres ..... 23
1.2.1.1 Shelterin complex ..... 26
1.2.1.2 Other telomere-associated proteins ..... 31
1.2.1.3 Telomere length measurement ..... 33
1.2.2 Regulation of telomere length ..... 35
1.2.2.1 Telomere replication and processing ..... 35
1.2.2.2 Telomere shortening and ageing ..... 37
1.2.2.3 Telomere elongation ..... 42
1.2.2.3.1 Telomerase inhibitors ..... 44
1.2.3 Telomere length inheritance ..... 45
1.2.4 Telomeres and MSI: role in cancer ..... 47
1.2.4.1 Cancer risk and genetic anticipation ..... 48
1.3 Colorectal cancer ..... 50
1.3.1 Hereditary colon cancer ..... 50
1.4 Hypothesis and main aims ..... 53
2 Materials and methods ..... 54
2.1 Materials ..... 54
2.2 Methods ..... 57
2.2.1 Cell culture ..... 57
2.2.1.1 Cell counts and population doubling ..... 57
2.2.1.2 Trypan blue exclusion ..... 58
2.2.1.3 $\beta$-galactosidase ..... 58
2.2.1.4 Transfection ..... 59
2.2.1.5 Telomerase inhibition ..... 60
2.2.1.6 Flow-cytometry analysis (FAC) ..... 61
2.2.2 Molecular biology ..... 61
2.2.2.1 DNA extraction from cell pellet ..... 61
2.2.2.2 DNA extraction from saliva ..... 62
2.2.2.3 Sanger sequencing ..... 63
2.2.2.4 DNA nanopore sequencing ..... 63
2.2.2.5 Dot blotting ..... 66
2.2.2.6 Western blotting ..... 67
2.2.2.7 Southern blotting ..... 68
2.2.2.8 Single telomere length analysis ..... 69
2.2.2.9 Telomere variant repeat PCR ..... 70
2.2.2.10 Telomerase repeat amplification protocol ..... 73
2.2.3 Statistical analyses ..... 73
3 MSH2 downregulation and telomere dynamics ..... 76
3.1 Background ..... 76
3.2 Aims ..... 79
3.3 Results ..... 80
3.3.1 Experiment design ..... 80
3.3.2 WI38 MSH2 $^{+/-}$clones growth ..... 85
3.3.3 MSH2 downregulation in WI38 clones ..... 90
3.3.4 Telomere shortening in WI38 MSH2 $^{+/-}$clones ..... 93
3.3.5 MSH2 deficiency, TSR and oxidative stress ..... 102
3.4 Discussion ..... 106
4 Telomere instability in $\mathrm{MSH}^{-/-}$cells ..... 109
4.1 Background ..... 109
4.2 Aims ..... 113
4.3 Results ..... 114
4.3.1 Mutant telomeres analysed by traditional TVR ..... 114
4.3.2 TVR by capillary electrophoresis ..... 119
4.3.3 Sequencing telomeres using the MinION ${ }^{\text {TM }}$ device ..... 130
4.4 Discussion ..... 148
5 Telomerase inhibition in MSH2 ${ }^{-/-}$cells ..... 152
5.1 Background ..... 152
5.2 Aims ..... 155
5.3 Results ..... 156
5.3.1 Partial inhibition of telomerase using BIBR1532 ..... 156
5.3.2 Cell cycle dynamics of cells treated with BIBR1532 ..... 160
5.3.3 Telomere dynamics and telomerase activity ..... 165
5.4 Discussion ..... 169
6 Telomeres in families ..... 174
6.1 Background ..... 174
6.2 Aims ..... 177
6.3 Results ..... 178
6.3.1 Recruitment and samples description ..... 178
6.3.2 Telomere length measurement and reproducibility ..... 183
6.3.3 Telomere shortening with age ..... 188
6.3.4 Telomere length inheritance ..... 197
6.4 Discussion ..... 211
7 Final remarks ..... 214
Appendix A Materials summary ..... I
Appendix B Primers summary ..... II
Appendix C Multiple alignment for OL A31 ..... III
Appendix D Multiple alignment for KUK A49 ..... XVII
Appendix E Multiple alignment for XpYp telomere ..... XXXI
Appendix F Students' contribution to the families study ..... L
Appendix G Technical reproducibility for control families ..... LII
Appendix H Technical reproducibility for LS families ..... LIII
Appendix I Control families pedigrees ..... LV
Appendix J Lynch syndrome families pedigrees ..... LXI
Bibliography ..... LXXIV

## List of Tables

1.1 Identity and functions of $E$. coli and eukaryotic homologous pro- teins involved in the MMR pathway ..... 12
1.2 Mouse lines with MMR defects and their relationship with fer- tility ..... 21
1.3 Summary of the main proteins at human telomeres ..... 32
1.4 Summary of the main methods to measure telomere length ..... 34
1.5 Review of the telomere shortening rates reported in the literature ..... 41
1.6 Main telomerase inhibitors ..... 44
2.1 Cell lines information ..... 54
2.2 Amplicons and primers for nanopore sequencing ..... 64
2.3 TVR methods compasion ..... 71
2.4 Conditions for the ABI 3100 genetic analyzer ..... 71
3.1 Summary of STELA results ..... 97
4.1 Mutation frequencies for LoVo ..... 129
4.2 Poretools stats output on the first sequencing experiment in the MinION ${ }^{\text {TM }}$ device ..... 136
4.3 Poretools stats output on the second sequencing experiment in the MinION ${ }^{\text {TM }}$ device ..... 140
4.4 Summary of the alignment statistics for the four libraries se- quenced ..... 147
5.1 Cell line information ..... 156
5.2 Telomere lengths and telomere shortening rates ..... 168
6.1 Features of the studied groups ..... 182
6.2 Mutation details per family ..... 184
6.3 Regression curves comparison ..... 195
6.4 Age-adjusted ( $\Delta$ ) telomere length and $\mathrm{P} / \mathrm{C}$ comparisons ..... 201
$6.5 \Delta$ telomere lengths and $\mathrm{P} / \mathrm{C}$ comparisons for LS families ..... 204

## List of Figures

1.1 DNA damage response ..... 4
1.2 MMR mechanism in E.coli ..... 7
1.3 MMR mechanism in eukaryotes ..... 13
1.4 Models proposed for the MMR mechanism ..... 16
1.5 Diagram showing replication slippage in a $(\mathrm{CA})_{4}$ microsatellite ..... 18
1.6 Diagram of the telomere structure and its possible conformations ..... 24
1.7 Shelterin proteins and interactions ..... 30
1.8 Telomere replication and processing ..... 36
1.9 Structure of telomerase ..... 43
2.1 Library preparation for minION ${ }^{\text {TM }}$ ..... 65
2.2 Diagram of STELA ..... 69
2.3 Diagram of TVR ..... 72
2.4 Diagram of TRAP assay ..... 75
3.1 WI38 growth curves under different oxygen conditions ..... 81
3.2 Puromycin dose-response curve for WI38 cells ..... 81
3.3 Plasmid and shRNA used for transfection ..... 83
3.4 Transfection efficiency ..... 84
3.5 Clones and population growth curves ..... 87
3.6 Growth curves for clones analysed for telomere length ..... 87
3.7 Senescence and viability ..... 89
3.8 Preliminary western blots ..... 91
3.9 MSH2 downregulation in WI38 clones ..... 92
3.10 STELA blots and scatter plots for WI38 clones ..... 94
3.11 Scatter plots showing allele specific lengths for XpYp ..... 95
3.12 Scatter plots for telomere length of WI38 transfected clones ..... 96
3.13 Relationship between TSR and the level of MSH2 ..... 99
3.14 Telomere shortening rates in WI38 transfected clones ..... 100
3.15 Telomere variant repeats for the WI38 cell line ..... 101
3.16 Effect of starting mTL on TSRs ..... 103
3.17 Telomere shortening rates under different oxygen conditions ..... 105
4.1 Single XpYp telomere molecules generated by smSTELA ..... 114
4.2 TVR optimisation conditions ..... 116
4.3 Primer labeling efficiency using dot blots ..... 117
4.4 Telomere variant maps from LoVo DNA ..... 118
4.5 Validation of TVR analysis by capillary electrophoresis ..... 120
4.6 Telomere variant map details of the progenitor XpYp A haplotype12
4.7 Details of the mutant XpYpA telomeres in LoVo ..... 125
4.8 Size of the smSTELA analysed by TVR ..... 126
4.9 Telomere variant repeat maps comparison for LoVo ..... 128
4.10 Graphical comparison of the mutation analyses in LoVo ..... 129
4.11 MinIONTM device description ..... 131
4.12 Pipeline for MinION ${ }^{\text {TM }}$ reads analysis ..... 133
4.13 Screen shots for Tel1 library sequencing run ..... 137
4.14 Screen shots for the control OL A31 library sequencing run ..... 138
4.15 Multiple sequence alignment for Tel1 library ..... 141
4.16 Screen shots of the channels panel for two flowcells ..... 143
4.17 Multiple sequence alignment for Tel2 library ..... 146
5.1 Telomerase downregulation in cell extracts ..... 157
5.2 Cell growth and death for short-term treatment with BIBR1532 ..... 158
5.3 Western blot for LoVo ..... 159
5.4 Growth dynamics of cancer cells treated with BIBR1532 ..... 160
5.5 Cell cycle description ..... 161
5.6 Proliferation index and cell division rate ..... 164
5.7 Cell cycle analysis by FACs ..... 164
5.8 TRAP assay on pellets from cancer cells treated with BIBR1532 ..... 167
5.9 Scatter plots for telomere length assayed by STELA ..... 168
6.1 DNA quality for control families ..... 179
6.2 Cohort description ..... 181
6.3 Technical reproducibility between students ..... 185
6.4 Correlation of XpYp and 12 telomere lengths ..... 186
6.5 Example of four STELA blots ..... 187
6.6 Telomere shortening with age in control individuals ..... 191
6.7 Telomere shortening with age in Lynch syndrome patients ..... 193
6.8 Regression curves comparison for telomere length with age ..... 194
6.9 Regression curves comparison between sexes ..... 196
6.10 Median telomere length in control families ..... 197
6.11 Median telomere length in Lynch syndrome families ..... 199
6.12 Age-adjusted telomere length ..... 206
6.13 Change in telomere length over generations ..... 207
6.14 Comparison of median telomere lengths within families ..... 210

|  | Abbreviations |
| :---: | :---: |
| ADP | Adenosine Diphosphate |
| ATP | Adenosine Triphosphate |
| BER | Base Excision Repair |
| bp | base pair |
| CEPH | Human Polymorphism Study Center |
| CIN | Chromosomal Instability |
| CRC | Colorectal Cancer |
| DAPI | 4¢,6-Diamidino-2-phenylindole dihydrochloride |
| DDR | DNA Damage Response |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic Acid |
| DN-hTERT | Dominant Negative human Telomerase Reverse Transcriptase |
| DSB | Double Strand Break |
| dsDNA | double stranded DNA |
| EDTA | Ethylenediaminetetraacetic acid |
| HNPCC | Hereditary Non-Polyposis Colorectal Cancer |
| HR | Homologous Recombination |
| FAC | Fluorescence-Activated Cell sorting |
| FAP | Familial Adenomatous Polyposis |
| FISH | Fluorescent In Situ Hybridisation |
| G4 | G-quadruplex |
| IDLs | Insertions/Deletions |
| kb | kilobases |
| LS | Lynch syndrome |
| LOH | Loss of Heterozygosity |
| mTL | median Telomere Length |
| MMR | Mismatch Repair |
| MSI | Microsatellite Instability |
| NER | Nucleotide Excision Repair |
| NHEJ | Non-homologous End Joining |
| nt | nucleotide |
| PBS | Phosphate-Buffered Saline |
| P/C | Parent/Children pair |
| PCNA | Proliferating Cell Nuclear Antigen |
| PCR | Polymerase Chain Reaction |
| PD | Population Doubling |
| PI | Proliferation Index |
| POL | DNA Polymerase |
| RFC | Replication Factor C |
| RNA | Ribonucleic Acid |
| ROS | Reactive Oxygen Species |
| RPA | Replication Protein A |
| shRNA | short hairpin Ribonucleic Acid |
| siRNA | small interfering Ribonucleic Acid |
| SSB | Single Stranded DNA Binding protein |
| ssDNA | single stranded DNA |
| STELA | Single Telomere Length Analysis |
| STR | Short Tandem Repeat |
| TERC | RNA component of telomerase |
| TERT | Telomerase Reverse Transcriptase |
| TRAP | Telomerase Repeat Amplification Protocol |
| TSR | Telomere Shortening Rate |
| TVR | Telomere Variant Repeat |
| UV | Ultraviolet light |

## 1 Introduction

### 1.1 DNA mismatch repair

### 1.1.1 Origin of mismatches in the DNA

The DNA mismatch repair (MMR) pathway detects and corrects mismatches in the DNA that can be either single-base, and there are eight different types, producing transition (A/C and G/T) or transversion (G/G, C/C, A/G, C/T, $\mathrm{A} / \mathrm{A}$ and $\mathrm{T} / \mathrm{T}$ ) mispairs or insertion/deletion loops (IDLs) that provoke gains or losses of nucleotides. The main causes involved in the formation of these mismatches are summarised in figure 1.1:

Physical or chemical damage of the DNA and its precursors. Some of the genotoxic agents that produce mismatches are reviewed in Hoeijmakers (2001), Lord \& Ashworth (2012) and Iyama \& Wilson (2013):

- Ionising radiation such as ultraviolet (UV) light produces covalent bonds that crosslink adjacent pyrimidines (cytosine and thymine) in the DNA strand.
- X-rays generate free radicals within the cell resulting in single-strand and double-strand breaks.
- Alkylating agents such as MNNG (1-methyl-3-nitro-1-nitrosoguanidine) that adds alkyl groups to the $\mathrm{O}^{6}$ of guanine and $\mathrm{O}^{4}$ of thymine producing $\mathrm{O}^{6}$-methylguanine and $\mathrm{O}^{4}$-methylthymine base modifications that can lead to transitions between AT and GT.
- Oxidative stress: elevated concentrations of ROS (Reactive Oxygen Species) at the extra and/or intra cellular lever produce oxidative DNA damage. ROS can interact with all four DNA bases and the sugar-phosphate backbone producing many oxidative products (reviewed in Dizdaroglu
(2012)). Recently, Cilli et al. (2015) showed that incorrect incorporation of ribonucleotides during replication, might facilitate formation of complex lesions involving oxidized nucleotides. The three main mechanisms that repair this type of lesion are: base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR), depending on the specific nucleotide modification (reviewed in Bridge et al. (2014)).

Replication. The DNA polymerases that synthesise nuclear DNA are: $\alpha$ and $\delta$ (mainly involved in the lagging strand replication) and $\epsilon$ (for the leading strand) (Lujan et al., 2012) and together with POL $\gamma$ (for mitochondrial DNA replication) are highly accurate polymerases, generating on average less than one base substitution or IDL for every 10,000 of correct incorporation events (McCulloch \& Kunkel, 2008). The error rate is different for each polymerase and is directly dependent on the sequence context and size, being higher in long repetitive regions than in short non-repetitive ones (Lujan et al., 2015). Because of the need to accurately replicate the genome, polymerases have developed mechanisms that favour the incorporation of the correct dNTP; besides, POL $\delta, \epsilon$ and $\gamma$ show 3 ' to 5 ' exonuclease activity which acting as proofreading control against incorrect base incorporation and decreasing the error rates compared to pol $\alpha$ that lacks this editing activity (McCulloch \& Kunkel, 2008). To increase replication fidelity, different post-replication repair pathways co-operate with polymerases: BER (Krokan et al., 2000), NER (Marteijn et al., 2014) and MMR (Modrich, 1991). It has been demonstrated in Saccharomyces cerevisiae that the efficiency of the MMR pathway to repair POL $\alpha, \delta$ and $\epsilon$ replication errors is different, correcting more efficiently potential deleterious errors to assure a high-fidelity replication (Lujan et al., 2012). The polymerase selectivity and the proofreading function of DNA polymerases result in an error rate of approximately $10^{-7} \mathrm{bp}$ per cycle of replication and including the repair mechanisms preventing that mutations arise during replication, the mutation rate in vivo has been estimated to be lower than $10^{-9} \mathrm{bp}$ per cell division (McCulloch \& Kunkel, 2008).

Genetic recombination. Two mechanisms occurring during genetic recombination affect the formation and processing of mispaired intermediates (Modrich \& Lahue, 1996):

- Gene conversion: This kind of homologous recombination involves the unidirectional transfer of genetic material from a "donor" sequence to a homologous "acceptor" with a very high sequence identity (Chen et al., 2007). Heteroduplex DNA structures are formed during the repair process, which are called Holliday junctions (Holliday, 2007).
- Homeologous recombination: This kind of recombination occurs between DNA sequences that are similar but not identical and generates mismatches in the DNA increasing genome instability (reviewed in Li (2008)).


### 1.1.2 MMR mechanism in prokaryotes

The first organism in which the MMR mechanism was studied was Salmonella pneumoniae and two genes, HexA (Balganesh \& Lacks, 1985) and HexB (Prats et al., 1985) were discovered. It was found that $\mathrm{Hex}^{-}\left(\mathrm{HexA}^{-}\right.$and $\left.H e x B^{-}\right)$mutants displayed a 5 to 30 -fold increase in spontaneous mutation rates, indicating that the Hex system played a role in mutation avoidance by correcting errors during replication (Tiraby \& Fox, 1973). Furthermore, transfection studies using $\lambda$ heteroduplexes demonstrated direct evidence for a mismatch repair system in Escherichia coli (reviewed in Claverys \& Lacks (1986)). Subsequently, the effect of mutator alleles on mismatch repair genes was studied and mutations in $U v r D, M u t H, M u t S$ or $M u t L$ were found to cause high mutability; in addition, DNA methylation was studied as a target for correct strand discrimination (Modrich, 1987).

### 1.1.2.1 MMR in E. coli

There are many pathways that repair DNA damage in $E$. coli but two main mechanisms can be distinguished on the basis of mispair specificity and size of the region that needs to be repaired (Modrich, 1991). The long-patch mismatch correction also called MutHLS mismatch repair pathway or E. coli


Figure 1.1: DNA damage response. Figure adapted from figure 1 in Hoeijmakers (2001). The main DNA damaging agents are shown on the left, the DNA damage that they produce is shown in the middle and the main DNA damage response pathway activated per DNA lesion is shown on the right. Abbreviations are: BER: base excision repair; NER: nucleotide excision repair; HR: homologous recombination; NHEJ: non-homologous end joining; MMR: mismatch repair.
methyl-directed system ((Modrich, 1991); (Fishel \& Kolodner, 1995); (Kolodner, 1996)) repairs single-base mispairs (except C-C), and the signal directing the strand-specificity is secondary and can be located very far away from the mismatch. The short-patch mismatch repair repairs up to 10 nucleotides IDLs, and the signal triggering the outcome of correction resides within the mismatch or very close to it. In this section, we will focus on the long-patch MMR, and the main proteins involved in it: MutS, MutL, MutH, DNA helicase II (MutU/UvrD), exonucleases (ExoI, ExoVII, ExoX and RecJ), SSB (Single Stranded DNA Binding protein), DNA polymerase III and DNA ligase. Figure 1.2 shows a diagram of the repair pathway that can be divided into three main steps:
a) Initiation: The mismatch is recognised by the MutS homodimer. The mutS gene was first isolated from Salmonella typhimurium by Pang et al. (1985) and it was shown to have homology to hexA. The crystal structure of the MutS dimer in E.coli was discovered by Lamers et al. (2000) and it was shown that the two MutS monomers have different conformations and only one of them recognizes the mismatch in an ATP dependent way (Junop et al., 2001). The asymmetry was described by Jiricny (2000) as "two praying hands with the thumbs coming close and the fingers slightly touching". Once the mismatch has been recognised, the MutH protein identifies the newly synthesized strand by binding to a hemi-methylated d(GATC) site (the daughter strand is transiently unmethylated) and because of its endonuclease function, it makes a nick in the unmodified strand, serving as a signal for direct excision repair (reviewed in Iyer et al. (2006)). MutH was first described by Welsh et al. (1987) in E.coli as a 28 KDa monomeric endonuclease dependent on $\mathrm{Mg}_{2}^{+}$that discriminated between strands. Evolutionarily, this protein was related to type-II restriction endonucleases (reviewed in Yang (2000)) but homologues of this protein have only been found in gram-negative bacteria suggesting that other mechanisms are used for strand specificity in other organisms. Junop et al. (2003) found that the active site of MutH in E. coli was composed of two separate structural domains and that the C-terminal $5^{\prime}$ residues of MutH affected DNA binding
and cleavage.
b) Excision: In a next step, the MutL homodimer binds to MutS and MutH and it has been demonstrated by Ban \& Yang (1998) that ATP-binding and not hydrolysis by MutL is essential for activating MutH. The mutL gene was also identified in S. typhimurium and E. coli (Pang et al., 1985) and it showed homology with the S. pneumoniae hexB gene (Prats et al., 1985) and (Claverys \& Lacks, 1986). The crystal structure of a $40-\mathrm{kDa}$ Nterminal fragment of $E$. coli MutL (LN40) was described in Ban \& Yang (1998) and Ban et al. (1999). They also demonstrated that MutL was an ATPase and that there was a structural similarity among MutL, DNA gyrase and Hsp90. The C-terminal domain of this protein included a $\beta$ binding motif that mediated a weak but specific interaction between MutS and MutL (Pillon et al., 2011). When the heterodimer between MutS and MutL is formed around the heteroduplex complex of DNA, the methyl-directed excision system is activated. Then, helicase II is recruited and it generates single-stranded DNA (ssDNA) around the nick made by MutH. This ssDNA is protected by single-strand binding proteins (SSB) from nuclease attack. MMR is bidirectional (Cooper et al., 1993), meaning that mistakes can be corrected from 3' or 5' direction and depending on the position of the strand break, different exonucleases are needed. ExoI and ExoX are 3' to 5' exonucleases while ExoVII and RecJ are 5' to 3' exonucleases. The first experiment that demonstrated a role of ExoI in MMR was in $S$. pombe (Szankasi \& Smith, 1995).
c) Resynthesis: It has been demonstrated that MutL physically interacts with the clamp subunits of DNA polymerase III (DiFrancesco et al., 1984) and this interaction is sufficient to support the repair synthesis step in E. coli. Finally, DNA ligase (Lahue et al., 1989) restores the DNA integrity by sealing the gap.


1. Hemi-methylated new strand with a mismatch
2. ATP-dependent MutS mismatch recognition

3. MutH binds to the methyl group
4. MutL binds to MutS and MutH.
MutH is activated by the binding involving ATP and nicks the new strand.

5. Exo 1 removes the mismatch. The MutS/MutL complex is released. RPA and Helicase 11 are recruited. Polymerase 111 corrects the mismatch.
6. DNA ligase 1 seals the nick
MutS MutL

Figure 1.2: MMR mechanism in E.coli. Schematic representation of the DNA mismatch repair mechanism in the prokaryote E. coli. The square at the bottom contains the names and figures used to represent each protein involved in this pathway and a brief summary of the actions occurring in each step is reported on the right part of the figure.

### 1.1.3 MMR mechanism in eukaryotes

The first study in eukaryotes identifying MutHLS homologous was done in Saccharomyces cerevisiae. Williamson et al. (1985) demonstrated that these genes were required for MMR and later, biochemical studies showed that MMR in eukaryotes depended on homologous proteins to the bacterial MutL and MutS (Reenan \& Kolodner, 1992). The MMR pathway is evolutionarily very conserved and some similarities such as substrate, bidirectionality and specificity are common between prokaryotes and eukaryotes. Nevertheless, the core proteins in humans, although homologous to the E.coli ones, are more numerous and the protein interactions more complex. In addition, the strand discrimination and excision mechanisms are different. Table 1.1 compares the bacterial and human proteins with their functions and interactions.

### 1.1.3.1 MMR mechanism in humans

Figure 1.3 shows a diagram of the MMR pathway in humans. Likewise in $E$. coli, the MMR pathway in humans can be divided into three main steps:
a) Initiation: The human homologous of the bacterial $M u t S$ is a group of proteins called MSH proteins (from Mut S Homologous) and the repair is initiated when either MutS $\alpha$ (MSH2-MSH6) which repairs single base-base and 1to 2 IDLs mismatches, or MutS $\beta$ (MSH2-MSH3) which repairs larger IDLs, heterodimers bind to a mismatch (Kunkel \& Erie, 2005). In humans, there are three different types of MSH proteins:

- hMSH2 ( $\sharp P 43246)$ : the gene was found to be near a locus implicated in hereditary non-polyposis colon cancer (HNPCC) (Peltomäki et al., 1993). The $h M S H 2$ gene is located in the human chromosome $2 \mathrm{p} 22-$ 21, has 16 exons and encodes a 105 kDa protein that selectively binds to DNA containing base-base mispairs and IDLs mismatches (reviewed in Kunkel \& Erie (2005)). It was shown by Fishel et al. (1994) and Umar et al. (1994) that cells containing mutant $h M S H 2$ genes exhibit microsatellite instability (MSI) and are defective in MMR (for more information refer to section 1.1.4). Repair activity can be restored in
these extracts by a protein preparation designated $h \mathrm{MutS} \alpha$ ((Drummond et al., 1995), (Umar et al., 1997) and (Watanabe et al., 2000)) consisting of a heterodimer of $h \mathrm{MSH} 2$ and $h \mathrm{MSH} 6$. Most of the germline and somatic mutations affecting $h M S H 2$ and other MMR genes were reviewed in Papadopoulos \& Lindblom (1997) and Peltomäki \& Vasen (1997). Missense mutations account for $17 \%$ of all identified alterations of this gene and Ollila et al. (2008) analysed some of these mutations located in the amino-terminal domain to demonstrate the pathogenicity of those mutations. Studies in vivo by Mastrocola \& Heinen (2010) generated human cell lines stably expressing some missense mutations and tested their effect on DNA repair and checkpoint response function. They found that those MSH2-deficient cells had lost their normal DNA checkpoint response to alkylating agents and also lost or decreased their MMR activity in vivo.
- hMSH6 ( $\sharp P 52701$ ): it was demonstrated by Drummond et al. (1995) and Palombo et al. (1995) that the $h$ MutS homologue was a heterodimer (called MutS $\alpha$ ) of approximately 100 KDa (corresponding to $h \mathrm{MSH} 2$ ) and $160-\mathrm{KDa}$ (firstly called GTBP; now known as $h \mathrm{MSH} 6$ ). This gene was localized by Papadopoulos et al. (1995) in 2p16 and they proposed that $h M S H 2$ and $h M S H 6$ might have been produced by duplication of a primordial mutS repair gene. Bowers et al. (1999) showed that mutations in a very conserved residue involved in binding to the mismatched base in S. cerevisiae conferred loss of MMR in vivo.
- hMSH3 ( $\# P 20585)$ : one mutS homologue located upstream of the human dihydrofolate reductase gene (DHFR) was identified by Fujii \& Shimada (1989) and located at 5q11.2-q13.2. In the mouse, a mutS homologue was also found and it was called Rep3 (mouse repair gene 3). When later on, in S. cerevisiae MSH3 was identified by New et al. (1993) as an open reading frame near the DHFR gene, a relationship was thought to exist between these three genes. Three years later, a frameshift mutation in $h M S H 3$ was observed in an endometrial carcinoma (Risinger et al., 1996) showing MSI that was partially
reduced when introducing human chromosome 5, so they proposed that $h M S H 3$ acts in the repair of some but not all mismatches.

All MutS homologous have highly conserved domains: one of them of approximately 150 amino acids forms an helix-turn-helix domain associated with an adenine nucleotide and a magnesium binding motif (Walker-A motif). It has been demonstrated that this domain has ATPase activity (Haber \& Walker, 1991). The MSH2-MSH6 complex has two different conformations: the ADP-bound form, called ON, that binds to mismatched nucleotides, and the ATP-bound form, called OFF, that, on the contrary, cannot bind to mismatches (Gradia et al., 1997). In addition, domains I and IV bind to DNA, domain V contains the dimerization interface and nucleotide-binding site and the N -terminal region contains a motif interacting with PCNA ( $\beta$-clamp homologue).
b) Excision: Equivalent to what happens in E. coli, the next complex that takes part in the repair pathway is a MutL homologue. In yeast, Prolla et al. (1994) identified two genes, MLH1 and PMS1 with homology to E. coli mutL and S. pneumoniae hexB. They also demonstrated that disruption of MLH1 resulted in high spontaneous mutation rates. Subsequently, three human mutL homologous were described by Papadopoulos et al. (1994):

- hMLH1 ( $\sharp P 40692$ ): is located in chromosome 3p21.3 and Hemminki et al. (1994) showed that loss of heterozygosity (LOH) of markers within or adjacent to this gene occur in some HNPCC tumours. Some mutations in $h M L H 1$ disrupt $h$ MLH1 protein forming complexes with $h E X O 1$ and PMS2 in vivo (Jäger et al., 2001). Somatic promoter hypermethylation has also been found in sporadic colon tumours.
- PMS1 ( $\sharp P 54277$ ): the name of PMS proteins comes from post-meiotic segregation as they were isolated by Williamson et al. (1985) after meiosis in S. cerevisiae. In humans, PMS1 has three isoforms (a, b and c) and mutations of this gene have been identified in some HNPCC patients (Nicolaides et al., 1994). PMS1 and hMLH1 form a heterodimer called MutL $\beta$ with unknown function in humans.
- PMS2 ( $\sharp P 54278)$ : is a $y$ PMS1 homologue endonuclease that has also been found mutated in HNPCC patients (Nicolaides et al., 1994). It forms a heterodimer with $h \mathrm{MLH} 1$ called MutL $\alpha$ (MLH1-PMS1 in yeast) coordinating events after mismatch binding by MutS homologue (reviewed in Hsieh \& Yamane (2008)).

The C-terminal domain of MutL in B. subtilis and E. coli specifically interacts with $\beta$ sliding clamp (Pillon et al., 2011), stimulating MutL endonuclease activity. PCNA is the human homologue of the $\beta$ subunit of $\boldsymbol{E}$. coli polymerase III and interactions between MutL and PCNA enhance polymerase $\delta$ processivity by enabling its sliding during the elongation process (O’Donnell et al., 1992) through its interaction with MutS $\alpha$, MutL $\alpha$ and $h$ EXO1 (Fiorentini et al., 1997). MutL $\alpha$ has a PCNA/RFC - dependent endonuclease activity that plays a critical role in 3 ' nick-directed MMR involving hEXO1 (Kadyrov et al., 2006). It is not clear yet how strand discrimination occurs in eukaryotes as no MutH homologue has been found; however, it was shown that small gaps of 4 to 18 nt might highlight the newly synthesised strand (Iams et al., 2002) and later, Ghodgaonkar et al. (2013) demonstrated that incorrect addition of ribonucleotides by DNA polymerases can initiate the MMR pathway.
c) Resynthesis: In eukaryotes, the MMR is bidirectional as in bacteria but in this case, $h$ EXO1 has both 5 ' and 3 ' exonucleolytic activities (Genschel et al., 2002) excluding the necessity for different exonucleases. PCNA is essential during 3 ' nick-directed MMR but not for 5’ nick-directed MMR (Li, 2008). It has been demonstrated that PCNA binds preferentially to the 3 ' terminus of a strand break whereas RFC (replication factor C) binds at the 5' terminus (Jiricny, 2006). RPA (replication protein A) stimulates hEXO1 activity in the presence of a MutS $\alpha$ complex bounded to a mismatch, stabilizes the single-stranded DNA and inhibits $h$ EXO1 activity in conjunction with $\operatorname{MutL} \alpha$ when the mismatch is removed. Finally, polymerase $\delta$ fills the gap and DNA ligase 1 seals the remaining nick.

| E.coli proteins | Function | Eukaryote homologous | Function |
| :---: | :---: | :---: | :---: |
| MutS (Su \& Modrich, 1986) | Recognises and binds mismatches | hMSH2-hMSH6 (MutS $\alpha$ ) (Fishel et al., 1994) hMSH2-hMSH3 (MutS $\beta$ ) (Fishel et al., 1994) | Recognises single base and 1-2 base IDL mismatches Recognises some single base IDLs and IDLs $\geq 2$ bases |
| MutL (Pang et al., 1985) | Intermediates the coordination of MMR | hMLH1-hPMS2 ( $\gamma$ PMS1) (MutL $\alpha$ ) (Hemminki et al., 1994); (Papadopoulos et al., 1994) and (Prolla et al., 1994) <br> MLH1-MLH2 ( $h$ PMS1) (MutL $\beta$ ) (Nicolaides et al., 1994) <br> MLH1-MLH3 (MutL $\gamma$ ) (Kolas et al., 2005) | Intermediates the coordination of MMR by binding MutS homologues <br> Unknown function in humans but it suppresses some IDL mutagenesis in yeast <br> Suppresses some IDL mutagenesis and participates in meiosis |
| MutH (Welsh et al., 1987) | Recognises hemimethylated GATC sites and nicks the unmethylated strand | None |  |
| $\gamma-\delta$ Complex | Loads $\beta$-clamp onto DNA | RFC complex | Loads PCNA <br> Modulates excision polarity |
| $\beta$-Clamp | Interacts with MutS <br> Enhances processivity of DNA pol III | PCNA (O'Donnell et al., 1992) | Interacts with MutS and MutL homologues <br> Recruits MMR proteins to mismatches <br> Increases MM binding specificity of MSH2-MSH6 <br> Participates in excision and probably in signalling <br> Participates in DNA repair synthesis <br> Participates in DNA re-synthesis |
| Helicase II | Loads onto DNA at nick by MutS and MutL <br> Unwinds DNA to allow excision of ssDNA | None |  |
| ExoI (Szankasi \& Smith, 1995), ExoX | Performs 3'-5' excision of ssDNA | $h \mathrm{EXOI}$ (Tishkoff et al., 1997a) | Performs excision of dsDNA |
| RecJ | Performs 5'-3' excision of ssDNA | 3' exo of pol $\delta$ | Performs excision of ssDNA |
| ExoVII | Performs 5'-3' and 3'-5' excision of ssDNA | 3 ' exo of pol $\epsilon$ | Synergistic mutator with Exo1 mutant |
| DNA pol III | Performs re-synthesis of DNA | DNA pol $\delta$ | Performs repair synthesis |
| SSB | Protects ssDNA from nucleases attack | RPA | Participates in excision and in DNA synthesis by stabilizing excision intermediates |
| DNA ligase | Seals nicks after DNA synthesis | DNA ligase I | Seals nicks after DNA synthesis |

Table 1.1: Identity and functions of $\boldsymbol{E}$. coli and eukaryotic homologous proteins involved in the MMR pathway. Adaptation from table 1 in Kunkel \& Erie (2005), table 1 in Hsieh
" $h$ " indicates human protein; " $y$ " indicates yeast protein; dsDNA: double stranded DNA; ssDNA: single stranded DNA; MM: mismatch


Figure 1.3: MMR mechanism in eukaryotes. Schematic representation of the DNA mismatch repair mechanism in humans. Adapted from figure 3 in Jiricny (2006) and figure 1 in Hsieh \& Yamane (2008). The square at the bottom contains the names and figures used to represent each protein involved in this pathway and a brief summary of the actions occurring in each step is reported on the right part of the figure.

### 1.1.3.2 Models for MMR mechanism

Three models have been proposed for MutS-MutL interaction during the MMR pathway. Two models are referred as "cis" or "moving" models because MutSMutL complexes are loaded at a mismatch site and then moved away from the site to search for the strand break that serves as strand discrimination signal. The third model is called "trans" or stationary model because the complex binds to the mismatch and only communicates with more distant signals by complex protein interactions. A diagram comparing the three models is shown in figure 1.4.
A. Hydrolysis-dependent translocation model:Allen et al. (1997) and Dao \& Modrich (1998) proposed that in the presence of MutL and ATP, the MutS protein covers the mismatch and the endonuclease activity of MutH (in E. coli) or MutL $\alpha$ (in humans) produce an incision, triggering the degradation of the newly synthesised strand. In this model, ATP hydrolysis is needed for bidirectional translocation of the MutS complex along the helix (reviewed in Kunkel \& Erie (2005)). DNA runs through the protein complex until MutS and MutL complexes reach a strand discrimination signal, forming a DNA loop (reviewed in Li (2008)). This model is not currently favored because there are not known molecular controls able to regulate ATP hydrolysis and/or ADP to ATP exchange in this process (Acharya et al., 2003).
B. Molecular switch model: Fishel (1998) and Ban \& Yang (1998) described this model based on the observation that hMSH heterodimeric complexes showed significant mismatch-dependent ATPase activity (reviewed in (Gradia et al., 1997) and (Gradia et al., 2000)). In this model, supported by the crystaline structure of MSH2 described by Obmolova et al. (2000), ATP binding to hMSH proteins produces the formation of a DNA sliding clamp. The term "sliding clamp" was first used to describe the $\beta$-subunit of $E$. coli DNA polymerase III and PCNA and refers to a protein complex that encircles DNA allowing it to pass through the hole
in its centre. MutS initially binds to mismatched DNA in an ADP bound state triggering a tridimensional change that allows an ADP to ATP exchange. This results in conformational changes allowing the clamp to slide through the DNA (reviewed in Li (2008)). It was suggested that the role of this switch is to control the timing of the downstream excision repair event Gradia et al. (1997). Later experiments refined this model proposing that MutS heterodimers act as a mismatch sensor and that a threshold number of localized ATP-bound MutS sliding clamps are required to initiate MMR (Acharya et al., 2003), introducing a redundancy into the model which is necessary to ensure that the MMR reaction can be restored from the last end-point if any dissociation happens during the process.
C. Transactivation model: Junop et al. (2001) proposed that MutS-MutL complexes can communicate with the mismatch and downstream proteins using protein-protein interactions and DNA bending. In this model, MutS heterodimers scan the DNA looking for a mismatch using the ATPase activity which acts as a proofreading enzyme to verify mismatch binding and authorizes the downstream excision (Junop et al., 2003). It was postulated that MutS heterodimers bind ATP after recognition of a putative mismatch and this binding is sufficient for recruitment of MutL $\alpha$ and downstream repair proteins (reviewed in Iyer et al. (2006)).

Those three models are suitable for the prokaryotes and eukaryotes MMR pathway and no consensus has been reached towards which one is the most likely. However, Haye \& Gammie (2015) showed that the MutS heterodimers are linked to the replisome possibly via their interaction with PCNA. They proposed a moving model in which MutS $\alpha$ or MutS $\beta$ load into the DNA as it is being replicated scanning for mismatches or small IDLs. Furthermore, Supek \& Lehner (2015) demonstrated that the MMR is more efficient repairing early replicating errors in euchromatic regions of the genome, suggesting that either the accessibility to the DNA and/or the coupling with the replication machinery are needed for correct repair.


Figure 1.4: Models proposed for the MMR mechanism. Adapted from figure 1 in Li (2008). The two "cis" or moving models are the TRANSLOCATION and MOLECULAR SWITCH and the "trans" is the STATIONARY model (reviewed in Acharya et al. (2003)).

### 1.1.4 MMR, MSI and cancer predisposition

Microsatellites are short tandem DNA repeat sequences (STRs) of 1 to 6 base pairs distributed throughout the genome which are prone to suffer from slippage errors during replication, producing microsatellite instability (MSI) (reviewed in Li et al. (2004)). This instability changes the length of the microsatellite, resulting in gain or loss of one or more repeat units and it can be used as diagnostic tool when comparing microsatellite profiles between tumours and normal tissue for the same patient (reviewed in Umar \& Kunkel (1996) and Shah et al. (2010)). See figure 1.5 for a schematic representation.

Knock-out mice for MMR genes showed overall higher mutation frequencies than MMR-proficient mice, with the highest mutuation rates in Mlh1-/and $M s h 2^{-/-}$mice (Hegan et al., 2006). Mutation rates in tandem repeats were assessed by transfection experiments using neomycin (neo) gene disrupted with a microsatellite. Spontaneous mutation rates of microsatellites in normal human cells have been proposed to vary between $3.1 \times 10^{-9}$ to $4.5 \times 10^{-9}$ mutations per cell per generation (Boyer \& Farber, 1998) whereas the mutation rate in MMR-defective cancer cells was reported to be approximately two orders of magnitude higher than in repair-proficient cancer cells by Boyer et al. (1995) and Hanford et al. (1998). Studies in tumour cell lines mutants for MSH2 (eg. LoVo or HEC59), MSH6 (HCT15), MLH1 (HCT116) or PMS2 (DU145) have shown MSI and high mutation rates in endogenous genes (Boyer et al., 1995) compared to MMR-proficient cancer cells (HT1080). It was demonstrated by Peltomäki et al. (1993) that microsatellites are highly unstable in most tumours from HNPCC (hereditary non-polyposis colon cancer, see section 1.3.1 for more details) patients and in some sporadic colon cancers (SCC). MSI is present in $15 \%$ of sporadic CRC (Liu et al., 1995) and in almost all hereditary CRC (2-5\% of all CRC) (Buecher et al., 2012). The risk of suffering cancer may be increased by the inheritance of different mutations called variants of uncertain significance (VUS) (Kantelinen et al., 2012). Inheriting different MMR VUS, which individually may not have any effect in the development of cancer, produce a synergistic effect contributing to the development of CRC. Diagnosis of HNPCC can be performed using 5 microsate-
llites markers (BAT26, BAT25, D5S346, D2S123 and D17S250) and tumours are classified as MSI-High (MSI-H) when two of more markers show MSI, MSI-Low (MSI-L) with one unstable marker or MSS (microsatellite stable) when no microsatellite loci are unstable compared to the normal tissue (reviewed in Vasen et al. (2007) and Lynch et al. (2009)). Interestingly, MSI has been shown to be reduced during treatment with aspirin in colorectal but not endometrial MMR-deficient tumours (Rüschoff et al., 1998) and thus, treatment with aspirin has been proposed as prophylactic treatment for families with an hereditary colon cancer predisposition.


Figure 1.5: Diagram showing replication slippage in a (CA) ${ }_{4}$ microsatellite. The template strand is shown in blue and the new strand in red. Black dots indicate Watson-Crick bounding. Adapted from figure 2 in Umar \& Kunkel (1996).

### 1.1.5 Other functions for MMR

Regardless its main role detecting and correcting replication errors, the MMR pathway controls several non-canonical functions reviewed in Iyer et al. (2006), Bak et al. (2014) and Crouse (2016).

Sensor for genetic damage: in normal (MMR-proficient) cells, treatment with cytotoxic agents such as 6 -thioguanine or MNNG leads to an accumulation in G2 phase of the cell cycle and activation of the p53 apoptosis pathway that protects cells against genetic damage acting as a mechanism to avoid cancer development. Treatment of MMR-deficient cells with the same cytotoxic agents, produces no accumulation in G2 phase nor an activation of the apoptosis pathway (reviewed in Iyer et al. (2006), Jiricny (2006) and Hsieh \& Yamane (2008)). This effect can be reversed by introducing the human chromosome 2 containing the wildtype of $h M S H 2$ and $h M S H 6$ into HEC59, LoVo or HCT15 ((Umar et al., 1997) and (Watanabe et al., 2000)). DNA methylators such as 5-fluoro-2'deoxyuridine (FdU) and cisplatin which are mutagenic (promoting nucleotide mispairs) and cytotoxic agents, have been used as antitumour drugs because they accumulate in proliferative cells promoting apoptosis. The effectiveness of this kind of chemotherapy is sometimes reduced in MMR-defective tumours as deficiency in the MMR pathway impedes to recognise the mismatch generated by FdU incorporation, resulting in poor response against cytotoxic agents in MMR-deficient patients ((Iyer et al., 2006) and (Liu et al., 2008)). Interestingly, it was found that selective inhibition of caspase 3 (a key protease involved in cellular degradation during apoptosis) in HCT116 ( $\mathrm{MLH1}^{-/-}$) cancer cells improved the response to chemotherapy with FdU (Flanagan et al., 2016), suggesting that in MMR-defective cancer cells, caspase 3 has a greater impact on cell proliferation than in apoptosis promotion.

Triplet repeat instability: some neurological, neurodegenerative and neuromuscular diseases such as myotonic dystrophy type 1 and Huntington's disease are caused by expansions of CTG and CAG trinucleotide re-
peats (TNRs) (reviewed in Iyer et al. (2015)). It was shown that MutL $\alpha$ and $\operatorname{MutS} \beta$ are required for the repair of CTG slippage replication errors in vitro (Panigrahi et al., 2012), but a transcriptional process might be involved in CAG tracks stabilization in vivo (Lin et al., 2006). Therefore, MSH2 deficiency in transgenic mice containing (CAG) ${ }_{\mathrm{n}}$ repeats was shown to prevent somatic instability (Manley et al., 1999) and to delay the age onset of the disease (Wheeler et al., 2003); additionally, MSH2 deficient transgenic mice containing (CTG) n repeats had a higher proportion of contractions in spermatogonia (Savouret et al., 2004) while depletion of MSH3 but not MSH6 reduced (CTG)n expansions in DM1 mice (van den Broek et al., 2002) and recently, a polymorphism in hMSH3 has been found to be associated to the variation in somatic instability in a group of DM1 patients (Morales et al., 2016). Furthermore, Msh2 ${ }^{-/-}$ transgenic mice for X-fragile syndrome showed an intergenerational reduction in CGG repeat expansions (Lokanga et al., 2014). Interestingly, it was found that some triplet repeat sequences, such as $(A T T)_{3}$, can form non-B-DNA structures that bind MMR proteins suppressing their activity what was suggested to be a natural cis-acting suppressor of the normal MSH2-dependent MMR machinery (Lujan et al., 2012).

Anti-recombination activity: some studies revealed that MutS and MutL were able to access mismatched base pairs within early strand transfer intermediates blocking the recombination process (reviewed in (Iyer et al., 2006)). In MMR-deficient cells, the homeologous recombination frequency was dramatically elevated, suggesting that MMR suppresses this kind of recombination. This might be illustrated by the fact that, in human cells, MutS $\alpha$ or MutL $\alpha$ deficiency increased the rate of gene duplication 50 to 100 -fold, what may contribute to cancer predisposition (Hsieh \& Yamane, 2008). The anti-recombination activity of MMR proteins (mainly MSH2 and MSH6) inhibiting interactions between homeologous sequences was likely to have a role promoting genome stability and preventing genome rearrangements ((de Wind et al., 1999) and (Harfe \& Jinks-Robertson, 2000a)).

Role in meiosis: MSH4-MSH5 and MLH1-MLH3 heterodimers constitute in mammals the meiotic MMR complex promoting meiotic recombination. Loss of any of those four proteins results in a decreased level of crossing-over events during meiosis (see Kirkpatrick (1999), Harfe \& Jinks-Robertson (2000a) and Clark et al. (2013)) having an effect on both, male and female fertility as summarised in table 1.2. Many human diseases are related with mutations in either hMSH4 or hMSH5 such as neoplasias (breast cancer, myeloma, lung cancer, ovarian cancer, glioma and colorectal cancer) or immune diseases (systemic lupus erythematosus, Kawasaki disease and type 1 diabetes) but also with reproductive disorders as azoospermia and premature ovarian failure, both caused by a defect on hMSH5 (reviewed in Clark et al. (2013)). Baker et al. (1995) demonstrated that male but not female mice mutants for Pms2 were infertile because they produced only abnormal spermatozoa but this effect was not present when the mutated gene was Msh2. Furthermore, the localization of MLH3 protein in mice, associated to repetitive regions -centromeres- in late pachytene spermatocytes, was upregulated in $\mathrm{Pms2}^{-/-}$but downregulated in Mlh1 $1^{-/-}$because in the absence of PMS2, MLH1 associates with MLH3 (Kolas et al., 2005).

| Genotype | Male Fertility | Female Fertility | References |
| :--- | :---: | :---: | :--- |
| $M l h 1^{-/-}$ | - | - | Baker et al. (1995) |
| $P m s 1^{-/-}$ | + | + | Prolla et al. (1998) |
| $P m s 2^{-/-}$ | - | + | Prolla et al. (1998) |
| $M l h 3^{-/-}$ | - | - | reviewed in Hsieh \& Yamane (2008) |
| $M s h 2^{-/-}$ | + | + | Kirkpatrick (1999) |
| $M s h 3^{-/-}$ | + | + | de Wind et al. (1999) |
| $M s h 4^{-/-}$ | - | - | reviewed in Neyton et al. (2004) |
| $M s h 5^{-/-}$ | - | - | reviewed in Clark et al. (2013) |
| $M s h 6^{-/-}$ | + | + | reviewed in Hsieh \& Yamane (2008) |
| $E x o I^{-/-}$ | - | - | reviewed in Hsieh \& Yamane (2008) |

Table 1.2: Mouse lines with MMR defects and their relationship with fertility. Adapted from table 2 in Hsieh \& Yamane (2008). " + " indicates fertile and "-" indicates infertile.

Generation of immunoglobulin diversity: immunoglobulin class switching and somatic hypermutation are mechanisms for increasing antibody diversity during antigen-stimulated B-cell differentiation (reviewed in Bak et al. (2014)). During this process G:U mispairs, that can be recognized and processed by the MMR pathway, are generated by activationinduced cytidine deaminase (AID) which deaminates cytosine residues to uracil (Iyer et al., 2006). It has been proposed that those residues are processed by error-prone DNA polymerases such as DNA polymerase $\eta$ resulting in a mutagenic process. It was demonstrated that $M l h 3^{-/-}$ mice had an increased mutation frequency in immunoglobulin (Ig) variable regions (reviewed in Hsieh \& Yamane (2008)) suggesting that this gene usually reduces the accumulation of mutations in this region. Experiments with knockout mice have demonstrated that the presence of MutS $\alpha$ promotes mutagenesis rather than prevents it because B cells from MMR-deficient mice display a low level of somatic hypermutation and reduced class switch recombination ((Jiricny, 2006) and (Li, 2008)).

### 1.2 Telomeres

### 1.2.1 Function and structure of human telomeres

Telomeres, the natural end of linear chromosomes in eukaryotes, are short tandem repeat (STRs) arrays (Blackburn, 1991). Human telomeres are organized as a double-stranded (dsDNA) tandem repeats that start after a chromosome specific subtelomeric region (Baird et al., 2003). The non-canonical repeats $\left((T G A G G G)_{n},(T C A G G G)_{n} \text { or (TTGGGG }\right)_{n}$ among others) are called telomere variant repeats (TVR) and are located closer to the subtelomeric region (Baird et al., 1995) and the canonical, (TTAGGG) ${ }_{n}$ repeats, that follow the TVR region are followed by a terminal 3' G-rich single-stranded (ssDNA) extension (called G-overhang) of approximately 100 nucleotides (reviewed in Garrido-Ramos (2012), Stewart et al. (2012) and Lu et al. (2012)). The number of repeats per telomere varies widely among species; while in humans it encloses $10-15 \mathrm{~kb}$ (in germline cells) in mouse and rats telomere length might be longer than 50 kb (Nandakumar \& Cech, 2013). Their G-rich composition makes them prone to form G-quadruplexes (G4) (Zahler et al., 1991) or alternative structures such as T-loops as shown in figure 1.6 (reviewed in Lipps \& Rhodes (2009) and Nandakumar \& Cech (2013)):

T-loops: are lariat-like structures resulting from the dsDNA invasion by ssDNA. They were discovered by Griffith et al. (1999) who also correlated the size of the t-loop circle with the length of the telomeric repeat array. It is not known if all telomeres have this structure or if it is formed during the cell cycle to allow replication, but it was proposed by Luke-Glaser et al. (2012) that there should be a balance between t-loop resolving and forming activities constituting a precise temporal regulation. This control should be mediated by rapid and reversible post-translational protein modifications; furthermore they proposed that the RTEL helicase is able to resolve t-loops when present. The technique called STORM, developed by Rust et al. (2006), allows to see t-loops with a better resolution and might resolve some unsolved questions about them.

G-quadruplexes: are G-quartets which are formed by the association of four guanines, in a square, planar structure and that is hydrogen-bounded by Hoogsteen-base pairing (non-Watson-Crick pairing that involves N7 atom of a purine) and stabilized by $\mathrm{K}^{+}$and $\mathrm{Na}^{+}$ions (reviewed in Lipps \& Rhodes (2009)). Interestingly, Zahler et al. (1991) demonstrated that these structures inhibit telomere elongation mediated by telomerase (see section 1.2.2.3). Taking this feature into account, later studies have designed small organic ligands that stabilize G-quadruplexes as anticancer therapies (an extensive review of those experiments together with the most recent studies in vivo about G-quadruplexes presence and function can be found in Lipps \& Rhodes (2009) and table 1.6).


Figure 1.6: Diagram of the telomere structure and its possible conformations. Purple line indicates the Grich strand and green line, the C-rich strand. The top part of the figure shows the telomere double stranded (dsDNA) and single stranded (ssDNA) regions, indicating the location of the canonical (TTAGGG ${ }_{\mathrm{n}}$ ) and non canonical repeat variants. The bottom part shows the two possible conformations in T-loops and G-quadruplexes as reviewed in Lipps \& Rhodes (2009) and Nandakumar \& Cech (2013).

The ends of linear chromosomes are targets for DNA damage response pathways and in addition, are strongly exposed to genotoxic stress and other mechanisms involved in telomeric loss; therefore, the two main problems of telomeres are:

End-protection problem: was first described by Muller \& Altenburg (1930) when chromosome end fusions were observed in the fruit fly (Drosophila melanogaster). Because of their structure, telomeres look like DNA double strand breaks (DSBs) that need to be repaired, activating the DNA damage response (DDR) pathway or cell cyle arrest (reviewed in de Lange (2009)):

- DNA repair: DSBs are potentially lethal for mammalian cells and they usually stimulate homologous recombination (HR), a DNA repair mechanism that uses sister chromatids as a repair template (Liang et al., 1998). When that template is not available, another repair mechanism is favored: non-homologous end joining (NHEJ). This mechanism is errorprone in contrast with HR, but both constitute the two major pathways for DSBs repair. The consequences of DSBs repair in telomeres are end-to-end chromosomal fusions that are highly unstable as demonstrated by Muller \& Altenburg (1930) because dicentric chromosomes are formed and they can initiate multiple bridge-breakage-fusion cycles.
- Cell cycle arrest: is activated as a response against DNA damage and the major regulators of this response are ATM (ataxia telangiectasia mutated kinase) and ATR (ataxia telangiectasia and Rad3 kinase) (for a review see Cimprich \& Cortez (2008)). Both kinases phosphorylate a group of substrates promoting cell-cycle arrest and DNA repair even though their mechanism of action and the effects of mutations in those genes are different. Activation of the ATM kinase pathway leads to up-regulation of p53 and induction of senescence (see section 1.2.2.2 for more details), as it can be demonstrated by the presence of telomere dysfunction-induced foci (Verdun \& Karlseder, 2007). It has been demonstrated that senescence occurs when the unprotected chromosome ends are recognized as DSBs because in those foci, DSB repair complexes co-localise with telomeres (Muraki et al., 2012).

End-replication problem: was predicted by James Watson in 1972 and one year later, Olovnikov proposed the "marginotomy theory of ageing" (re-
viewed in Verdun \& Karlseder (2007)). In this theory, the "telogenes" (ends of DNA molecules) were ramdomly shortened in each mitosis cycle, providing a mechanism for ageing that agreed with the Hayflick limit (Hayflick \& Moorhead, 1961). The end-replication problem is caused by the inability of DNA polymerases to regenerate the RNA primer of the most distal Okazaki fragment, resulting in the loss of some nucleotides each time a cell divides. In $S$. cerevisiae the telomere erosion rate in normal cells has been reported as 3 to 4 bp per generation while the erosion rate in human cells is estimated between 50 to $200 \mathrm{bp} /$ cell division ((Baird, 2008), (Palm \& de Lange, 2008) and (Muraki et al., 2012)) and in mice $\approx 7 \mathrm{~kb} / \mathrm{cell}$ division (Vera et al., 2012), indicating the great variability in telomere shortening rates among eukaryotes. The progressive loss of telomeric DNA has been suggested to be a molecular clock reflecting the number of cell divisions that have undergone and contributing as a signal for the entrance into senescence (Harley et al., 1990) and (Chadeneau et al., 1995). It has been estimated that human sperm telomeres are about 10 to 14 kb long, whereas somatic cell telomeres are only 5 to 10 kb long, reflecting the number of divisions that have undergone in somatic cells (Pickett et al., 2011).

### 1.2.1.1 Shelterin complex

The shelterin complex caps the most distal part of chromosomes, preventing genomic instability (reviewed in Palm \& de Lange (2008), O'Sullivan \& Karlseder (2010) and Stewart et al. (2012)) as it a) enables cells to distinguish their natural chromosome ends from dsDNA or ssDNA strand breaks, b) represses DNA repair reactions by compacting telomeric chromatin (Bandaria et al., 2016), c) regulates telomerase-based telomere maintenance (Xin et al., 2008), d) is proposed to be implicated in the formation of t-loops and e) controls the synthesis of telomeric DNA by telomerase (de Lange, 2005). It is composed of six telomere-specific proteins (TRF1, TRF2, TIN2, Rap1, TPP1 and POT1) that can be found together in a single complex or forming different subcomplexes (see below for further information of each protein). The common characteristics of the shelterin complex compared with other
telomere-associated proteins are that a) it is abundant at chromosome ends but does not accumulate elsewhere, b) it is present at telomeres throughout the cell cycle and c) its function is limited to telomeres (de Lange, 2005).

- TRF1 (Telomeric repeat-binding factor 1) ( $\sharp$ P54274): first described by Zhong et al. (1992) is a 50 kDa protein binding to dsDNA and preferentially recognising the telomeric repeat sequence present at mammalian chromosome ends (as it does not bind well to other tandem repeat arrays such as TAGGG, TTTAGGG, TTTTAGGG, TTGGGG and TTAGGC). It is proposed that TRF1 negatively regulates telomere length in telomerase-positive human cells as its interaction with PINX1, an inhibitor of telomerase, decreases telomere length (reviewed in Xin et al. (2008)). Besides, diminished TRF1 loading onto telomeres has been found in ageing endothelial cells with very short telomeres (Hohensinner et al., 2016) suggesting that deprotection of telomeres by loss of TRF1 might increase telomere attrition. TRF1 also promotes replication throughout the telomere duplex as its removal from telomeres leads to replication fork stalling, defects in the telomeric tract packaging and fusions of sister telomeres (Stewart et al., 2012). In addition, TRF1 binding to the C-terminus of ATM constitutes a negative regulatory feedback that diminishes the ability of TRF1 to interact with DNA, delaying mitotic entry and apoptosis (reviewed in Xin et al. (2008)).
- TRF2 (Telomeric repeat-binding factor 2) ( $\ddagger$ Q15554): described by Bilaud et al. (1996) is known to homodimerize and it has been proposed to stabilize t-loop formation and to protect the telomeres by binding dsDNA, since: a) it preferentially binds to the end of a dsDNA TTAGGG repeat array if it contains a G-strand overhang of at least 6 nt , b) the multimeric binding mode of TRF2 introduces positive supercoils that promote unwinding and strand invasion and c) its ability to bind to Holliday junctions (reviewed in de Lange (2002) and Palm \& de Lange (2008)). Therefore, TRF2 primarily acts in chromosome end protection by promoting topological changes in telomeric DNA, t-loop and chromatin assembly and the suppression of ATM-dependent DNA damage response and

NHEJ (reviewed in O'Sullivan \& Karlseder (2010)). $\operatorname{Trf2}{ }^{-/-}$mice, which die early in embryogenesis (de Lange, 2005), demonstrated that TRF2 is an essential protein. The dominant negative TRF2 $\Delta \mathrm{B} \Delta \mathrm{M}$ is a truncated version that forms inactive heterodimers with the endogenous protein and effectively removes TRF2 from telomeres (van Steensel et al., 1998), resulting in end-to-end chromosome fusions. Conversely, overexpression of TRF2 has been found to increase telomere shortening (Karlseder et al., 2002) in an oxidative stress-independent way (Richter et al., 2007) that relies on the formation of telomeric ultrafine anaphase bridges (Nera et al., 2015).

- TIN2 (TRF1-interacting nuclear protein 2) ( $\ddagger$ Q9BSI4): was identified by its interaction with TRF1, in a two-hybrid screen, and it was localized at telomeres (Kim et al., 1999). Although both TRFH domains in TRF1 and TRF2 are very similar, they cannot heterodimerize, so TIN2 is necessary to bring them together. Additionally, TIN2 binds TPP1, thus providing a bridge between the shelterin components that bind to dsDNA and ssDNA (see figure 1.7). The subcomplex formed by TRF1-TIN2-TRF2-TPP1 stabilizes the shelterin complex but whether other subcomplexes of shelterin coexist in cells remains unknown (reviewed in Nandakumar \& Cech (2013)). However, it is known that TIN2 contains mitochondria targeting signals and knock down of TIN2 by siRNA enhanced oxygen consumption and mitochondrial ATP synthesis (Lu et al., 2012), suggesting that TIN2 may play a role regulating aerobic glycolysis and oxidative phosphorylation. Furthermore, mutations in TIN2 have been found in some dyskeratosis congenita (DC) patients and knock-in mice carrying this mutation had accelerated telomere shortening over generations (Frescas \& de Lange, 2014).
- Rap1 (Repressor and activator protein 1) ( $\sharp Q 9 N Y B 0)$ : is a 399 amino acids protein of approximately 47 KDa that interacts with TRF2 by its Cterminal domain (Li et al., 2000) and shares its structure with Rap1p (its yeast homologue) although in contrast to it, human Rap1 only has one Myb-like domain, lacking the one that directly binds to telomeric DNA (Palm \& de Lange, 2008). The Rap1 C-terminus also contains a putative nuclear localization signal (NLS) (de Lange, 2005) and inhibition of Rap1 by dominant-negative expression leads to elongated telomeres and loss of telomere heterogeneity (Li et al., 2000).
- TPP1 (POT1 and TIN2-interacting protein) ( $\sharp Q 96 A P 0): ~ s i m i l a r l y ~$ to TIN2, TPP1 also acts as a bridge between different shelterin components. This protein has a C terminus of 60 aminoacids that binds to the N-terminal of TIN2 and a central 100-amino-acid region which binds to the C terminus of POT1 (Liu et al., 2004). The N terminus of TPP1 has an OB-fold (oligonucleotid/oligosaccharide-binding) domain that interacts with telomerase raising the possibility that TPP1 is involved in the recruitment or regulation of telomerase (Palm \& de Lange, 2008). TPP1 increases the affinity of POT1 for DNA by 10 -fold and it also recruits telomerase in vivo (reviewed in Hwang et al. (2012)). Furthermore, depletion of TPP1 by shRNA resulted in inappropriate telomere elongation (Liu et al., 2004) which is a phenotype shown when POT1 is diminished. TPP1 is also required to localize POT1 to the telomeres due to the lack of a NLS in POT1 (reviewed in Stewart et al. (2012)). The interactions between TIN2-TPP1 and TPP1-POT1 occur in both the cytoplasm and the nucleus, contrasting to TIN2-TRF2 interaction (that exclusively takes place in the nucleus), what suggests that there should be a nuclear export signal (NES) in TPP1 that has been localized next to the POT1-recruitment domain on TPP1 (Xin et al., 2008).
- POT1 (Protection of telomeres protein 1) ( $\ddagger$ Q9NUX5): was identified by Baumann \& Cech (2001) due to its homology with the DNA-binding domain of $\operatorname{TEBP} \alpha$, a ciliate telomere protein. The two OB-fold domains in POT1 allow it to specifically bind the ssDNA overhang, preventing
inappropriate ATM activation (reviewed in Palm \& de Lange (2008)). It was suggested that POT1 binds to G4 in a sequential manner due to the two OB folds and that two POT1 monomers are needed to unfold the G4 (reviewed in Hwang et al. (2012)); furthermore, the POT1-TPP1 complex shows a highly dynamic sliding movement on the telomeric overhang, inducing continuous unfolding and refolding of the G4 (Hwang et al., 2012) and determining telomere length by competing with telomerase for access to the G-overhang (reviewed in O'Sullivan \& Karlseder (2010)).


Figure 1.7: Shelterin proteins and interactions. A shows the secondary structure of the six shelterin proteins indicating the domains in different colours. The TRFH homology domain acts as a dimerization domain in both TRF1 and TRF2. The SANT/Myb DNA-binding domain in C-terminal of TRF1 and TRF2, but not in Rap1, binds to double stranded DNA (dsDNA). The D/E rich domain (acidic aminoacids) in TRF1 is a binding site for tankyrase 1 and 2. The GAR domain (Gly/Arg-rich) is a basic domain with a sequence-independent affinity for DNA junctions, including Holliday junctions. The FxLxP motif in TIN2 binds to the TRFH domain of TRF1 and the N-terminus of TIN2 binds not only to TRF2 but also to TPP1 (Palm \& de Lange, 2008). The BRCT domain in the N-terminal of Rap1 is a central Myb-type helix-turn-helix motif and the acidic C-terminal binds to TRF2 (reviewed in Palm \& de Lange (2008)). Dotted coloured lines indicate protein-protein interaction colour-codded based on the domains involved in it. Dashed black lines indicate interactions of some of these domains with either DNA or telomerase (adapted from figure 2 in Palm \& de Lange (2008)). B 3D representation (not based on crystalographyc structure) of the shelterin complex. C shows hypothetical location of the shelterin complex at telomeres (adapted from figure 2 in de Lange (2005) and figure 1 in Stewart et al. (2012)).

### 1.2.1.2 Other telomere-associated proteins

There are lots of proteins with a function at telomeres and most of them do not form part of the shelterin complex. These non-shelterin factors are typically much less abundant at telomeres than the shelterin proteins and some are only transiently associated, whereas the shelterin complex is present at telomeres throughout the cell cycle (reviewed in de Lange (2005)). Besides, most of these non-shelterin proteins have non-telomeric functions, mainly related with DNA repair (Ku70/80, XPF/ERCC1, Apollo, the MRN complex, RAD51D, tankyrases 1 and 2), DNA damage signaling (the MRN complex), DNA replication (ORC, RecQ helicases) or chromatin structure (HP1 proteins) (reviewed in Palm \& de Lange (2008)). The interactions between all those proteins and the shelterin complex are very complicated and an example of the main protein-protein interactions on telomeres is shown on figure 1 in Xin et al. (2008) showing the telomere interactome. Here, I will enumerate the main proteins and their main interaction and function at telomeres:

- Tankyrases (TRF1-interacting, ankyrin-related ADP-ribose polymerase): Smith et al. (1998) discovered tankyrase 1 and 2 in a yeast two-hybrid screen with TRF1 and showed that their central domain (ankyrin) interacted with the acidic domain of TRF1 but not with TRF2 because the homologous domain is basic. In addition, its function in vitro was described as a negative regulator of TRF1 because ADP-ribosylation inhibited the ability of TRF1 to bind to telomeric DNA. Later, in vivo experiments demonstrated that tankyrase 1 is highly expressed in telomerasenegative cells such as WI38 and that overexpression of this protein in the nucleus of telomerase-positive cells induced loss of TRF1 and promoted telomere lengthening (Cook et al., 2002).
- WRN and BLM: those are RecQ helicases responsible for Werner syndrome and Bloom syndrome respectively, when mutated. Both were demonstrated to bind to TRF2 by ChIP analysis (Opresko et al., 2002) and although the function for BLM on telomeres is still unknown, WRN is likely to resolve G4 (reviewed in de Lange (2005)).
- MRN Complex: Mre11, Rad50 and Nbs1 form this complex that recognizes and accumulates in DNA DSB sites having multiple functions for maintenance of the genome integrity. Its presence at telomeres increases when they are deprotected and its role is usually related to HR and NHEJ (reviewed in de Lange (2005)).
- ERCC1/XPF: is a complex that functions as a structure-specific endonuclease that cuts DNA duplexes adjacent to a 3 ' ssDNA. Mutations in those genes have been found in mice but not in humans and ERCC1/XPFdeficient cells develop a novel telomere phenotype called telomeric DNAcontaining double minute chromosomes (for a review see de Lange (2005)). It was shown that this complex associates with TRF2, is present at telomeres, and is involved in overhang processing (Zhu et al., 2003)

|  | Proteins | Main functions on telomeres | Main Interactions |
| :---: | :---: | :---: | :---: |
| Shelterin complex | TRF1 | Negative regulation of TL | dsDNA, TRF2, TIN2, ATM, BLM, PINX1, Tankyrase |
|  | TRF2 | Promotes T-loop formation <br> Prevents recognition of dsDNA breaks Inhibits DSB repair | dsDNA, TRF1, TIN2, ATM/ATR, Rap1, WRN, PARP1, Apollo Nbs1/Mre11/Rad50, ERCC1/XPF |
|  | TIN2 | Bridge between dsDNA and ssDNA binding proteins | TRF1, TRF2 and TPP1 |
|  | Rap1 | Improves affinity and selectivity of TRF2 for dsDNA | TRF2 |
|  | TPP1 | Increases POT1 affinity for ssDNA Regulates TL by recruiting telomerase | POT1 and telomerase |
|  | POT1 | Prevents ssDNA degradation <br> Maintains G4 <br> Regulates TL by competing with telomerase | ssDNA, TPP1 |
| Nonshelterin proteins | Tankyrases | Regulates TL through inhibition of TRF1 | TRF1 |
|  | WRN | Resolves G4 | TRF2, BLM, p53, Ku70/80, PCNA |
|  | BLM | Unknown function | TRF2, WRN, ATM, Rad51, MLH1 |
|  | Mre11, <br> Rad50, <br> Nbs1 | DNA damage sensor | TRF2, ATM |
|  | ERCC1/XPF | Overhang processing | TRF2 |

Table 1.3: Summary of the main proteins at human telomeres. Proteins belonging to the shelterin complex are shown at the top and non-shelterin proteins at the bottom. The main telomeric function and interactions are shown. Abbreviations are: TL: telomere length; dsDNA: double-stranded DNA; ssDNA: single-stranded DNA; G4: G-quadruplexes. For more information about the proteins refer to section 1.2.1.1 for shelterins and 1.2.1.2 for nonshelterins.

### 1.2.1.3 Telomere length measurement

The knowledge of telomere dynamics mainly arises as a result of telomere length (TL) measurements under different conditions. It was shown that TL varies depending on the DNA extraction method (Cunningham et al., 2013), the method used to measure TL and even between laboratories (Martin-Ruiz et al., 2015). Therefore, it is important to explain here the main available and well established methods for TL measurement including their advantages and disadvantages (see table 1.4 for a summary).

Terminal Restriction Fragments (TRF) was the first technique used to measure telomere length (Harley et al., 1990). In this method, genomic DNA is digested by frequent cutter restriction enzymes that do not recognize telomeric and subtelomeric sequences. Then, the undigested subtelomeric and telomeric regions are resolved by size in an agarose gel, Southern blotted and hybridized. TL is measured by the labeled DNA smear that can be sized using a DNA ladder comparison; although extensively used, considerations such as probe intensity (greater for longer sequences) have to be taken into account before measurement. The next method to size TL (qPCR) was developed by Cawthon (2002) and lately modified for multiplexing purposes (Cawthon, 2009). This PCR method measures TL as a ratio between the telomere (T) and a single copy gene (S) amplification. More recently, an adaptation of the original qPCR method has been described (SCT-pqPCR) to measure TL in single cells after a pre-amplification step for T and S template. Baird et al. (2003) developed single telomere length analysis (STELA), a PCR based method that uses a subtelomeric-specific primer and a linker-specific primer that is ligated to the 5 ' of the telomere using the 3 ' G-overhang as a template. In a small-pool PCR, the subtelomeric primer and a reverse primer sharing the 5 ' sequence with the linker-primer amplify single telomeres (as shown in figure 2.2 in chapter 2) at sub-visible level. Southern blot and hybridization with a labeled-telomere probe is needed before individual telomeres are sized by a DNA size ladder comparison (a more detailed description of the method is shown in chapter 2 section 2.2.2.8). Finally, two methods based on fluorescence in situ hybridisation (FISH) are available: Q-FISH (Lansdorp et al.,
1996) and Flow-FISH (Rufer et al., 1998) in which a fluorescently labeledPNA probe is hybridised to the telomeres and detected and quantified by software. In both cases, viable cells are needed but while in Q-FISH cell spreads (metaphase or interphase) are made in slides and detection is performed using a microscope, in Flow-FISH, labeled-cells are separated and analysed using flow cytometry (Baerlocher et al., 2006).

| Method | Measures | Advantages | Disadvantages |
| :---: | :---: | :---: | :---: |
| TRF <br> Kimura et al. (2010) | Average TL | Well established method <br> Mean TL for total cell population <br> Not specialized equipment required <br> Optional overhang measurement | Large amounts of DNA ( $0.5-5 \mu \mathrm{~g}$ ) <br> Time-consuming (3-5 days) <br> 1 kb resolution (overestimation of TL) <br> TVR sequence and length might affect DNA digestion |
| qPCR <br> Cawthon (2002) <br> MMQ-PCR <br> Cawthon (2009) | Average TL | Adapted for high throughput Small amount of DNA (20ng) | TL referenced to standard single copy gene <br> Reference standards lacking <br> Need for replicate testing and standard samples <br> Should be used only for diploid and karyotypic stable cells |
| SCT-pqPCR <br> Wang et al. (2013) | Single-cell average TL | High throughput <br> Small amount of DNA <br> Compares TL between single cells <br> Lower CV than qPCR | Based on qPCR |
| STELA <br> Baird et al. (2003) | Chromosomespecific TL | Measures single telomeres <br> Resolution 0.1 kb <br> High sensitivity <br> Small amounts of DNA ( 250 pg ) <br> Allows to size critically short telomeres | Time-consuming (3-5 days) <br> Need of unique primers for each telomere <br> Lack of primers for all ends Underestimation of TL due to poor detection of long telomeres |
| Q-FISH <br> Lansdorp et al. (1996) | Average TL and chromosomespecific (metaphase) | Detects single telomeres <br> Low cell numbers <br> Strong signal and easily quantifiable <br> Allows to see telomere fusion events and telomere-free ends | For chromosome spreads cell viability is needed <br> TL cannot be measured in senescent cells or slow dividing cells High intra-assay variation Requires digital microscopy |
| Flow-FISH <br> Rufer et al. (1998) | Average TL | Adapted for high throughput Allows to compare between cell subpopulations <br> First TL method validated for clinical purposes | Limited to blood samples Calibration needed to measure very small telomeres (weak fluorescence) |

Table 1.4: Summary of the main methods to measure telomere length. Table adapted from table 1 in Aubert et al. (2012b), table 1 in Ozturk et al. (2014) and tables 1 and 2 in Montpetit et al. (2015). Abbreviations are: TRF: terminal restriction fragments; TL: telomere length; TVR: telomere variant repeat; qPCR: quantitative PCR; MMQ-PCR: monochrome multiples qPCR; SCT-pqPCR: single-cell telomere length qPCR; CV: coeficient of variation; STELA: single telomere length analysis; Q-FISH: quantitative fluorescence in situ hybridisation.

### 1.2.2 Regulation of telomere length

### 1.2.2.1 Telomere replication and processing

Replication of telomeric DNA is a multi-step process that involves not only the passage of a replication fork along the DNA duplex but also processing of the DNA terminus to generate the 3 ' overhang (Stewart et al., 2012). The formation of the 3 ' overhang is important due to its two main functions: a) serving as a template for telomere elongation by telomerase (see section 1.2.2.3 and b) allowing the formation of T-loop as a protective structure against telomere degradation (see section 1.2.1) (reviewed in Bonetti et al. (2013)). Replication of human telomeres, like other regions in the human genome, occurs in a semiconservative way and due to the end replication problem, the lagging strand automatically loses a few nucleotides, generating a small 3 ' overhang. However, 3 ' overhangs are observed at both, lagging and leading strands, suggesting that a more complex, other than replication, process controls its formation (reviewed in Arnoult \& Karlseder (2015)). In fact, Chow et al. (2012) demonstrated that generation of the overhangs occurs in a different manner (timely and mechanistically wise) for lagging and leading strands: while the former has mature 3 ' overhangs rapidly after replication, the latter needs a longer processing time that includes 5' C-strand resection.

It has been proposed that regulation of TL is controlled by a homeostatic pathway that is achieved through a negative feedback loop established by the shelterin complex ((Cong et al., 2002) and (Bandaria et al., 2016)). In long telomeres, telomerase (see section 1.2.2.3) might be inhibited because lots of shelterin complexes bind to telomeres whereas for short telomeres, there are only few shelterin complexes, what cannot inhibit telomerase (de Lange, 2005). All proteins forming the shelterin complex seem to be implicated in this negative regulation but POT1, which is the only able to bind ssDNA, is the one proposed to block telomerase through competing for the 3 ' overhang with telomerase.


### 1.2.2.2 Telomere shortening and ageing

Telomere length can be maintained by the telomerase enzyme that is active in germline, immortal cells and some stages during early embryogenesis, between the morula and blastocyst stages (see section 1.2.2.3). However, in most tissues telomeres shorten with every cell division (Harley et al., 1990) and consequently, during human life. Hence, telomeres have been proposed to be a molecular clock that controls ageing (Bodnar et al., 1998). The natural state of cell proliferation arrest due to accumulation of DNA damage response (DDR) signals (Fumagalli et al., 2014) including short or damaged telomeres, is called senescence and it is a protective mechanism preventing the accumulation of DNA lesions and hence considered as a tumour suppressor mechanism (reviewed in Suram \& Herbig (2014) and Muñoz-Espín \& Serrano (2014)). Interestingly, telomeres have been found shorter but less variable in length in healthy older than younger people (Halaschek-Wiener et al., 2008), suggesting that disease resistance and healthy ageing might be related with an optimal telomere length maintenance.

Analysing single cells in S. cerevisiae, Xu et al. (2015) showed that there are two very different types of lineages leading to senescence; on the one hand, the commonly called "senescent cells" (type A cells) that suffer a unique severe event triggering cell proliferation arrest and on the other hand, the "cryptic cells" (type B cells) that suffer some reversible cell cycle arrest earlier in time before senescence takes place. Those two types of lineages correlate with the two distinct types of senescence that can be distinguished in human cells (Nelson et al., 2014):

Replicative Senescence (RS) is the consequence of several consecutive rounds of cell divisions and is linked to the "Hayflick limit", telomere shortening being the main trigger (this would be equivalent to type A cells in S. cerevisiae), although many other factors have also been described (reviewed in Zeman \& Cimprich (2014)). Interestingly, protection of telomeres and not necessarily their relative length has been described to be involved in RS, considering that over expression of TRF2 prevented end-to-end fusions of critically short telomeres thus, delaying RS (Karlseder et al.,
2002). In addition, as telomere length at senescence can vary between cell types (von Zglinicki et al., 2000a) it has been proposed that RS is directly related with age more than with telomere length itself (Serra \& von Zglinicki, 2002). However, telomere shortening rates with age have been shown to vary between different primary cell types, human populations and age ranges (see table 1.5). In addition, cells with mutant p53 can escape RS and will undergo crisis triggered by telomere fusions produced as a consequence of the uninterrupted telomere shortening (Hayashi et al., 2015).

Induced Senescence (IS) or premature senescence on the contrary, is not necessarily linked to cell divisions since it has been shown that primary fibroblasts can undergo senescence after a long period of quiescence (Marthandan et al., 2014). IS can be initiated either by DNA damage or by oncogene activation such as Ras, Raf or MEK resulting in a cancer-preventing mechanism that can be reverted. Wei et al. (2001) demonstrated that in normal human fibroblasts, the growth arrest that precludes senescence occurs in the p53-p21-RB pathway at the level of p53. Premature senescence is characterized by the formation of senescen-ce-associated heterochromatin foci ((Schulz \& Tyler, 2005) and (Zhang et al., 2007)) that seems to be absent in RS (Kosar et al., 2011), as well as the presence of different protein expression profiles compared to RS. The higher expression of $\beta$ galactosidase activity is shared for both RS and IS (Gary \& Kindell, 2005), making it a standard assay for detecting senescence although new membrane markers such as DEP1 or BM2G have recently being proposed as another suitable detection method in aged cells (Althubiti et al., 2014).

| Author | Method | Cohort size [P] (T) | Type of study | Age range | Age $\pm$ SD | Telomere shortening rate $\pm$ error (bp/year) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Hastie et al. (1990) | TRF | 47 (blood, PBL) | cross-sectional | 20-85 | 52.5 | 33 |
| Allsopp et al. (1992) | TRF | 31 (skin and fibroblasts) | cross-sectional | 0-93 | 46.5 | $15 \pm 6$ (donor age) |
| Vaziri et al. (1993) | TRF | 119 (blood, PBL) | cross-sectional | 0-107 | 53.5 | $40 \pm 3.6$ (women) and $50 \pm 4.2$ (men) |
| Slagboom et al. (1994) | TRF | 123 [MZ and DZ] (blood, PBMC) | cross-sectional | 2-95 | 48.5 | 31.0 |
| Iwama et al. (1998) | TRF | 30 (blood, PBMC) | cross-sectional | 4-39 | 21.5 | 84 |
| Iwama et al. (1998) | TRF | 50 (blood, PBMC) | cross-sectional | 40-95 | 67.5 | 41 |
| Frenck et al. (1998) | TRF | 85 (blood, PBL) | cross-sectional | 0-82 | 41 | $12 \pm 1.2$ |
| Mondello et al. (1999) | TRF | 26 (blood, PBMC) | cross-sectional | 26-104 | 65 | 30 |
|  |  |  | cross-sectional | 0-90 | 45 | 39 (granulocytes) and 59 (lymphocytes) |
|  | Flow |  | bisegmented | 0-1 | 0.5 | 3052 (granulocytes) and 1088 (lymphocytes) |
| Rufer et al. (1999) | FISH | >500 [36 MD-DZ] (blood, PBL) | subset | 0-4 | 2 | 483 (granulocytes) and 457(lymphocytes) |
|  |  |  | subset | 4-90 | 47 | 36 (granulocytes) and 52 (lymphocytes) |
| Na | TRF | 124 (normal + CRC tissue) | cross-sectional | 29-97 | 62 | 44 (normal mucosa) |
| Na | TRF | 124 (normal + CRC tissue) | cross-sectional | 29-97 | 62 | 50 (cancer tissue) |
| von Zglinicki et al. (2000b) | TRF | 37 (blood, PBMC) | cross-sectional | 18-98 | 58 | 20 |
| Friedrich et al. (2000) | TRF | 9 (blood, PBL) |  | 73-95 | 84 | 50 |
| Friedrich et al. (2000) | TRF | 7 (skin) | cross-sectional | 73-91 | 82 | 78 |
| Benetos et al. (2001) | TRF | 193 (blood, WBC) | cross-sectional | - | $56 \pm 11$ | 36 (women) and 38 (men) |
| Takubo et al. (2002) | TRF | 191 (PM liver) <br> 137 (PM renal cortex) | cross-sectional | 0-104 | 52 | $\begin{aligned} & \hline 60 \\ & 46 \end{aligned}$ |
| Martens et al. (2002) | $\begin{aligned} & \text { Flow- } \\ & \text { FISH } \end{aligned}$ | 51 (blood, PBL) | cross-sectional | 0-92 | 49.0 | $15 \pm 3$ (T cells) and $33 \pm 3$ (B cells) |
| Nakamura et al. (2002) | TRF | 52 (PM epidermis) <br> 48 (PM lingual epithelium) | cross-sectional | 0-101 | 50.5 | $\begin{aligned} & 37.5 \pm 1.5 \\ & 31 \pm 1 \end{aligned}$ |
| Cawthon et al. (2003) | $\begin{aligned} & \text { qPCR } \\ & +\mathrm{TRF} \end{aligned}$ | 143 (blood, PBMC) | cross-sectional | 60-97 | 78.5 | 14 |
| Nawrot et al. (2004) | TRF | 327 [P/C pairs] (blood, WBC) | cross-sectional | - | - | 19 (women) and 24 (men) |
| Gardner et al. (2005) | TRF | 70 [50 whites, 20 blacks] (blood, WBC) | longitudinal | 21-44 | 32.5 | 31.3 (tel.attrition in 64; tel. gain in 5; no change in 1) |
| Martin-Ruiz et al. (2005) | qPCR | 598 (blood, PBMC) | longitudinal | 85-101 | 93 | $5 \pm 13$ (no significant) |
| Nordfjäll et al. (2005) | qPCR | 132 [P/C pairs] (blood, PBMC) | cross-sectional | 32-86 | 59 | 16 (women) and 25 (men) |
|  |  |  | cross-sectional | 30-80 | 55 | 31 (women) and 14 (men) |
| Unryn et al. (2005) | TRF | 125 (blood, PBMC) | subset | 30-49 | 39.5 | 51.3 |
|  |  |  | subset | 50-80 | 65 | 19.8 |
| Valdes et al. (2005) | TRF | 1122 [women] (blood,WBC) | cross-sectional | 18-76 | $47.8 \pm 12.1$ | 27 |
| Aviv et al. (2006) | TRF | 1517 [women TwinsUK] (blood, WBC) | cross-sectional | 18-79 | $48 \pm 12.7$ | $20.5 \pm 1.1$ |
|  |  |  | subset | 18-50 | 34 | 26 |
|  |  |  | subset | 51-79 | 65 | 25 |
| Cherkas et al. (2006) | TRF | 1552 [women TwinsUK] (blood, WBC) | cross-sectional | 18-75 | 46.5 | $19.8 \pm 1.7$ |


| Author | Method | Cohort size [P] (T) | Type of study | Age range | Age $\pm$ SD | Telomere shortening rate $\pm$ error (bp/year) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| De Meyer et al. (2007) | TRF | 2433 [Asklepios study population] (bood, PBMC) | cross-sectional | 35-55 | $46.2 \pm 5.9$ | 20.7 (women) and 31.0 (men) |
| Njajou et al. (2007) | $\begin{aligned} & \text { qPCR } \\ & + \text { +TRF } \end{aligned}$ | 907 [Amish families; P/C pairs] (blood, PBL) | cross-sectional | 18-92 | 55 | 40 |
| Steer et al. (2007) | TRF | 1151 (blood, WBC) | cross-sectional |  | 48.2 | 38 |
| Bataille et al. (2007) | TRF | 1897 (blood, PBL) | cross-sectional | 18-79 | 46.0 | $27 \pm 1.5$ |
| Barwell et al. (2007) | TRF | 1696 [Twin Res.Unit, St. Thomas' Hospital] (blood, WBC) | cross-sectional | 45-77 | 56.3 | 22 |
| Cherkas et al. (2008) | TRF | 2401 [Twins UK] (blood, PBL) | cross-sectional | 18-81 | $48.8 \pm 12.9$ | $21 \pm 1.3$ |
| Richards et al. (2008) | TRF | 1319 [Twins UK, $92 \%$ women] (blood, PBL) | cross-sectional | - | $40.9 \pm 12.5$ | 18.5 |
| Hunt et al. (2008) | $\begin{aligned} & \mathrm{TRF}+ \\ & \text { qPCR } \end{aligned}$ | 1395 [NHLBI and BHS cohorts] (blood, PBL) | cross-sectional | 19-93 | 56 | 20 (whites) and 29 (blacks) |
| Aviv et al. (2009) | TRF | 635 [BHS; 450 whites and 185 black] (blood, PBL) | longitudinal | 20-40(+6) | 31.4 | 37.8 (whites) and 40.7 (blacks) <br> (tel. attrition in 561; tel. gain in 68; no change in 6) |
| Ehrlenbach et al.(2009) | $\begin{aligned} & \text { qPCR + } \\ & \text { TRF } \end{aligned}$ | 510 [Bruneck Study] (blood, PMBC) | longitudinal | 45-85 (+10) | 58 | 45.5 (tel. attrition in 84\%; tel. gain in 15.9\%) |
|  |  |  | subset | < 51 | - | 39 |
|  |  |  | subset | 52-58 | 55 | 32 |
|  |  |  | subset | 59-66 | 62.5 | 48 |
|  |  |  | subset | $>67$ | - | 28 |
| Epel et al. (2009) | qPCR | 236 [MacArthur Health Aging Study] (blood, PBL) | longitudinal | 70-79 (+2.5) | 73.7 | Tel. attrition in 30\%; tel. gain in $24 \%$; no change in $46 \%$ |
| Farzaneh-Far et al. (2010) | $\begin{aligned} & \text { qPCR } \\ & + \text { TRF } \end{aligned}$ | 608 (blood, PBL) | longitudinal | - | $65 \pm 11$ | 42 (tel. attrition in 276; tel. gain in 140; no change in 192) |
| Houben et al. (2011) | qPCR | 75 [Zutphen Elderly Study] (blood, WBC) | longitudinal | 73-91 (+7) | 78-83 | 40.2 (tel. attrition in 84\%; tel. gain in 16\%) |
| Chen et al. (2011) | TRF | 271 [BHS] (blood, PBL) | cross-sectional | S1:19.9-41.5 | 30.7 | 24.6 |
|  |  |  |  | S2:25.7-46.3 | 36 | 25.4 |
|  |  |  |  | S3:31.5-49.9 | 40.7 | 23.6 |
|  |  |  | longitudinal | S1-S2 (+5.8) | - | 31.4 (tel. attrition in 86\%; tel. gain in 14\%) |
|  |  |  |  | S2-S3 (+6.6) | - | 33.5 (tel. attrition in 86\%; tel. gain in 14\%) |
|  |  |  |  | S1-S3 (+12.4) | - | 32.2 (tel. attrition in 98.5\%; tel. gain in 1.5\%) |
| Svenson et al. (2011) | $\begin{aligned} & \text { TRF + } \\ & \text { STELA } \end{aligned}$ | 50 (blood,PBL) | longitudinal | - | - | tel. attrition in 50\%; tel. gain in 50\% |
| Fitzpatrick et al. (2011) | TRF | 1136 [CHS] (blood, PBL) | longitudinal | $>65$ | - | 26 |
| Daniali et al. (2013) | TRF | 87 (blood PBL, muscle, skin, fat) | cross-sectional | 19-77 | 44 | 26 (leukocytes), 24 (muscle), 23 (skin) and 25 (fat) |



[^0] Heart Study; STELA: Single Telomere Length Analysis; CHS: Cardiovascular Health Study; GW: Gestational Week; LSADT: Longitudinal Study of Aging Danish Twins.

### 1.2.2.3 Telomere elongation

Very short telomeres have been proposed to activate TERRA (telomeric repeat containing RNA) transcription thus promoting recruitment and nucleation of telomerase (Cusanelli et al., 2013). Telomerase is a eukaryotic ribonucleoprotein complex that adds TTAGGG repeats to the human telomeres by using an RNA template ((Greider \& Blackburn, 1985) and reviewed in Blackburn et al. (1989)) and it is mainly active in germline and immortal cells.

The two main components of the catalytic core of telomerase are: TERT (also called TRT or Est2), that has homology with the catalytic motifs of the reverse transcriptase family and which main function on telomeres is to synthesise telomere DNA repeats, and TERC (also called TR or TLC1), that constitutes an RNA template of $\approx 450 \mathrm{bp}$ with sequence CUAACCCUAAC (reviewed in Greider (1996)). See figure 1.9 for further details on telomerase structure. In addition to its telomeric lengthening function, telomerase also protects against double strand breaks in a telomere synthesis independent manner (Fleisig et al., 2016), and has other non-canonical functions reviewed in Low \& Tergaonkar (2013).

As a mechanism compensating for the DNA erosion during replication, after C-strand resection, telomere elongation by telomerase produces extension of the 3 ' overhang and later, the C-strand can be filled in by Pol $\alpha$-primase (reviewed in Stewart et al. (2012), Nandakumar \& Cech (2013) and Bonetti et al. (2013)). Another non-telomerase-based mechanism for telomere maintenance called ALT (alternative lengthening of telomeres) has been described (Bryan et al., 1997) and is almost exclusively found in some tumours such as soft tissue sarcomas. Its mechanism involves recombination using sister chromatids as template and it has been detected in anomalous situations such as human cancer cells (10-15\%) or telomerase-null mouse cell lines. Nevertheless, some evidence suggest that this mechanism might also occur in normal cells under certain situations (reviewed in de Lange (2005) and Podlevsky \& Chen (2012)).

A


B


Figure 1.9: Structure of telomerase. Adapted from Kim et al. (2008) and Podlevsky \& Chen (2012). A represents TERT motifs and domains. TEN domain: is a basic amino-terminal domain containing DNA-binding domains in a single-stranded way, it is involved in TERT localization to the nucleolus and it contains RNA interacting domain 1 (RID1) with affinity by the pseudoknot domain in TERC and a DAT motif implicated in telomerase recruitment. TRBD domain: is implicated in RNA recognition and binding (Rouda \& Skordalakes, 2007) and the interaction between TRBD and TERC through the CR4/5 domain is required for the proper assembly and enzymatic activity of the holoenzyme, both in vitro and in vivo. RT domain: is the reverse transcriptase domain containing seven conserved motifs ( $1,2, \mathrm{~A}, \mathrm{~B}, \mathrm{C}, \mathrm{D}, \mathrm{E}$ ), which tertiary structure remains a right hand, with two domains (1 and 2) binding to the incoming nucleotides and five ( $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}$ and E ) constituting the catalytic site (reviewed in de Lange (2005)). CTE domain: shares functionality with the HIV RT C-terminus binding to the RNA template/DNA primer duplex. Mutations in this domain affect telomerase processivity and a DAT domain has also been identified, suggesting that this region may also be involved in telomerase recruitment to the telomere. B represents TERC motifs and domains. Template/pseudoknot: contains a triple helix and a template sequence that can be divided into a 5' region encoding for telomeric DNA repeats and a 3 ' region annealing to the DNA primer after template translocation (reviewed in Podlevsky \& Chen (2012)). CR4/5: is the other domain required for enzymatic activity. It is distal to the pseudoknot domain and composed of a three-way junction of helices. H/ACA: is located at the 3' end of TERC and contains two stem-loops separated by the box H/ACA moieties acting as binding sites for dyskerin, NOP10, NHP2 and GAR1. There is another domain, located in the 3' stem-loop, called Cajal body localization (CAB box) that serves as a signal for TERC localization in Cajal bodies, where the RNA component of the telomerase is packaged (reviewed in O'Sullivan \& Karlseder (2010)). The H/ACA domain is essential for TERC biogenesis (reviewed in Podlevsky \& Chen (2012)).

### 1.2.2.3.1 Telomerase inhibitors as therapeutic agents

Telomerase is active in $80-90 \%$ of human cancers ((Kim et al., 1994) and (Shay \& Bacchetti, 1997)), thus, it has been extensively studied as a therapeutic approach against cancer. Different targets have been proposed and table 1.6 summarises the most important ones (Phatak \& Burger, 2007):

|  | Inhibitor | Description and effects |
| :---: | :---: | :---: |
| hTERC <br> inhibition | PNAs: Peptide Nucleic Acids (Fleisig et al., 2016) | First telomerase inhibitors used <br> Trigger telomere loss <br> Very high cytotoxic effects <br> The drug GRN163 has been used to reduce tumour volume in xenograft mice and later in phase II clinical trials |
|  | As-ODNs: Antisense oligodeoxynucleotides Jiang et al. (2003) | Short ssDNA sequences complementary to TERC <br> Initiate antiviral pathway mediated by RNase-L, breaking down TERC <br> Telomere loss has not being studied High cytotoxic effects |
|  | GDEPT: Gene-directed enzyme prodrug therapy Plumb et al. (2001) | TERC promoter activity is 300 -fold higher in cancer cells than in normal cells <br> Combination of the bacterial NTR gene (under TERC promoter) and the pro-drug CB1954 sensitizes cancer cells |
| hTERT <br> inhibition | Nucleoside analogs (Olivero et al., 1997) | AZT has being tested in cell lines showing inhibition of telomerase, telomere shortening and cytotoxic effects High drug concentrations are required |
|  | BIBR1532 <br> (Damm et al., 2001) | High concentrations induce acute cytotoxic effecs The most specific synthetic hTERT inhibitor to date Produced telomere shortening in several cancer cell lines Requires lag-time (dependent on TL) before cell death or senescence |
|  | DN-hTERT <br> (Dominant negative TERT) (Hahn et al., 1999) and (Zhang et al., 1999) | Is reversible <br> Is non-cytotoxic <br> Produces telomere shortening and proliferative defects ending in growth arrest <br> Mechanism: cytoplasmic exportation of endogenous TERT and degradation via ubiquitination <br> Might affect non-telomeric functions |
| Telomere targeting agents (TTA) | G4 ligands <br> (Hurley, 2002) | Examples: BRACO19 or telomestatin <br> Produce disruption of the T-loop <br> Effectively inhibit telomerase and POT1 binding <br> Reduces telomere elongation <br> Some cytotoxic effects are observed <br> Used in xenograft mice to reduce tumour volume |
|  | T-oligos <br> Wojdyla et al. (2014) | Homologous to the 3 ' overhang <br> May mediate DDR via p53/p73 pathway <br> Inhibit proliferation by inducing apoptosis and/or senescence No telomere length effect |

Table 1.6: Main telomerase inhibitors. Table adapted from tables 1 and 2 in Rankin et al. (2008) and data found in Phatak \& Burger (2007). Abbreviations are: ssDNA: single stranded DNA; NTR: bacterial nitroreductase; AZT: azydothymidine; TL: telomere length. DDR: DNA damage response.

### 1.2.3 Telomere length inheritance

It has largely been discussed how telomeres are inherited since a high heritability ( $\mathrm{h}^{2}$ ) index has been described (70-78\%) based on twin studies ((Slagboom et al., 1994), (Graakjaer et al., 2004) and (Broer et al., 2013)) and on linkage disequilibrium ( $81.9 \pm 11.8 \%$ ) for $12 q$ telomere (Vasa-Nicotera et al., 2005). However, other studies have suggested lower $h^{2}$ values ( 36 and 44\%) based on quantitative-trait linkage analysis (Njajou et al., 2007) and statistical analysis (Andrew et al., 2006). Moreover, the distinct contribution of the progenitors to their offspring's telomere length is still under discussion (Eisenberg, 2014), resulting in more studies suggesting that fathers have a stronger impact on their offspring's TL ((Nawrot et al., 2004), (Nordfjäll et al., 2005), (Unryn et al., 2005), (De Meyer et al., 2007), (Nordfjäll et al., 2010) and (Holohan et al., 2015)) than mothers (Broer et al., 2013). Those studies concluded that, as a positive correlation between sperm TL and donor-age has been found ranging from 71 bp/year (Allsopp et al., 1992) to $135 \mathrm{bp} /$ year (Baird et al., 2006), older parents might transmit longer telomeres to their offspring ((Kimura et al., 2008) and (Eisenberg et al., 2012)). Nevertheless, other studies have suggested that this "lengthening" of sperm TL with age might be due to a decrease of the TL over generations, that therefore, distorts the data ((Holohan et al., 2015) and (Stindl, 2016)). It is important here to describe telomere dynamics in the germ line and during gametogenesis, to understand the parental contribution to their offspring's TL:

Oogenesis: Liu \& Li (2010) and Ozturk et al. (2014) reviewed several studies that have analysed the presence of telomerase in the ovary. It has been described that telomerase is necessary to maintain TL in the ovarian surface epitelium (stem cells). Therefore, telomerase activity would contribute to folliculogenesis and ovulation as short telomeres in the ovary and lack of telomerase in granulosa cells was related to ovarian insufficiency and female infertility (Keefe \& Liu, 2009). During oocyte maturation, telomerase activity is reduced and an important telomere shortening (from $11.41 \pm 0.81 \mathrm{~kb}$ to $8.9 \pm 0.86 \mathrm{~kb}$ ) is observed despite
the absence of cell divisions (Turner \& Hartshorne, 2013). Furthermore, telomerase is known to be activated by estrogen (either by a direct activation of TERT gene transcription or by an indirect mechanism involving MYC). Other factors, such as oxidative stress or accumulation of unrepaired DNA damage during meiotic arrest, might contribute to this telomere shortening observed in humans, as it has been demonstrated to occur in mice (Yamada-Fukunaga et al., 2013).

Spermiogenesis: telomerase is mainly active in primary spermatocytes and it disappears in mature differentiated spermatozoa (reviewed in Ozturk (2015)) suggesting that mature sperm TL is defined very early during maturation. Furthermore, short telomeres in sperm have been related to male infertility (Ferlin et al., 2013) although telomerase activity levels were not different between fertile and infertile testis (Fujisawa et al., 1998), suggesting that its presence in infertile testis does not restore TL. Finally, it is also important to consider that the male pronuclei has shorter telomeres than the female one (Turner \& Hartshorne, 2013), what might be related with the greater number of cell divisions occurring in the former compared to the later. Interestingly, Reig-Viader et al. (2014) found higher levels of TERRA colocalising with telomeres in spermatocytes I than in oocytes I along all phases of gametogenesis, what suggests that TL in sperm cells might be related to the abundance of these transcripts at telomeres.

A minimum TL in the germline is likely to be needed to produce a competent embryo (reviewed in Kalmbach et al. (2014)) and a model is proposed in which after fertilisation, telomere lengthening occurs during the earliest pre-implantation development (morula-blastocyst), correlating with the high levels of telomerase activity detected in this stage ((Wright et al., 1996) and (Wright et al., 2001)). This telomere elongation was proposed to be needed to reach a set point after which the zygote genome can be activated (Schaetzlein et al., 2004). Then, telomeres would be maintained (possibly due to ALT in addition to telomerase) and telomere shortening might start together with adult cell differentiation and the progressive loss of telomerase activity that
was found to be different among distinct tissues (Wright et al., 1996). In adulthood, telomere shortening rates have been described to be two fold higher in longitudinal versus cross-sectional studies (reviewed in Unryn et al. (2005) and see table 1.5 for more details) possibly due to changes in the TL over generations; however, Verhulst et al. (2013) suggested that baseline TL might have an impact overestimating shortening rates. To overcome this issue, although more expensive and time consuming, longitudinal studies should be preferred versus cross-sectional as they represent a direct observation of a biological process (during the follow-up time) that might be masked when performing cross-sectional analysis.

### 1.2.4 Telomeres and MSI: role in cancer

Development of human cancer has been associated with genetic instability and in the case of colorectal cancer (see section 1.3), at least two forms of this instability have been described (reviewed in Fodde et al. (2001b) and Centelles (2012)). On the one hand, microsatellite instability (MSI) caused by MMR defects (present in $10-15 \%$ of sporadic colorectal cancers (SCC) and $\approx 90 \%$ of familial cancer), results in a mutator phenotype characterized by changes in the number of repetitive sequences of microsatellites (see figure 1.5). On the other hand, chromosomal instability (CIN) (in $65-70 \%$ of SCC) produces an accelerated rate of gains or losses of whole or large portions of chromosomes resulting in karyotypic variability from cell to cell and as a consequence, aneuplody, subchromosomal genomic amplifications and a high frequency of loss of heterozygosity (LOH) (Pino \& Chung, 2010).

Telomere length and telomerase activity are not only implicated in the control of the proliferative capacity of normal cells but also in malignant progression. Takagi et al. (2000) showed that high MSI (MSI-H) colorectal tumours showed shortened telomeres suggesting that MSI may be associated with telomere shortening in colorectal carcinogenesis despite the presence of telomerase (Hiyama et al., 1995). Furthermore, Pickett et al. (2004) and Mendez-Bermudez \& Royle (2011) showed a higher telomere mutation frequency (defined by the appearance of mutations that changed the sequence
of the assayable proximal variant repeat region of the telomere) in MMRdefective tumours (specially $\mathrm{MSH2}^{+/-}$) compared to MMR-proficient and hypothesised that a tendency to losses of repeats might be triggering an increase rate of telomere shortening, contributing to the profile of short telomeres in colorectal tumours. In addition, the telomeric varian repeat (CTAGGG) n was found to be associated with an exceptional mutation rate in the male germline that was only transmitted to the offspring via male but not female germline (Mendez-Bermudez et al., 2009). Cells with dysfunctional, very short uncapped telomeres, can a) enter replicative senescence as a tumour suppressing mechanism (reviewed in Martínez \& Blasco (2011), Muraki et al. (2012) and Suram \& Herbig (2014)) or b) continue dividing and escape cellular crisis by a p53 mutation and enter breakage-fusion-bridge cycles that can continue for multiple cell generations leading to a dramatic genome reorganization (Pino \& Chung, 2010). Interestingly, MSH2 deficiency in Terc ${ }^{-/-}$mice was shown to abolish the protecting mechanism of dysfunctional telomeres, thus increasing lifespam but without an increase in telomere recombination (Martinez et al., 2009). Besides, Jones et al. (2014) showed in an MLH1 ${ }^{-/-}$cancer cell line that ligase 3 (specially the BRCT domain) was required for cells with very short telomeres to escape cellular crisis.

### 1.2.4.1 Cancer risk and genetic anticipation

Genetic anticipation is the phenomenon whereby the symptoms of a genetic disorder become apparent and gradually more severe at an earlier age on successive generations (see more information in the MedGen database). The mechanism for anticipation in many neurological disorders has been related with triplet repeat expansions (reviewed in Iyer et al. (2015)) suggesting that repetitive structures such as telomeres might as well being involved in the earlier age onset of certain types of tumours. Dyskeratosis congenita (DC), an inherited predisposition for a bone marrow failure that might result in malignancy, is caused by germline mutations in genes directly affecting telomerase (reviewed in Gu et al. (2009)); hence, it is not surprising that DC patients have very short telomeres and that genetic anticipation has been related with
progressive telomere shortening over the generations (Vulliamy et al., 2004). Interestingly, telomeres in DC patients were predicted to have a reduced telomere attrition rate with age than controls (cross-sectional study in Alter et al. (2007)) due to its very short length although, when some of those patients were followed up over time, telomere seemed to shorten at similar rate than controls for some and even to lengthen for others (Alter et al., 2012). Furthermore, despite shorter age-adjusted telomere lengths in aplastic anemia (a subtype of bone marrow failure) compared to controls, patients with longer telomeres responded better to immunosuppressive therapy (Brümmendorf et al., 2001). Progressive shortening of telomeres has also been correlated with disease progression in chronic myelogenous leukaemia (reviewed in Bronner et al. (1994). Martinez-Delgado et al. (2011) showed that telomeres shortened in following generations of hereditary breast cancer but later, Pooley et al. (2014) found no association between telomere length and cancer risk in BRCA mutation carriers, although telomeres were found longer in those compared to their mutation non-carrier relatives. However, there might still be a relationship between telomere stability and cancer risk in hereditary breast cancer as Benitez-Buelga et al. (2016) showed that a specific mutation in OGG1, that repairs oxidative damage and has been described to maintain telomere stability (Wang et al., 2010), was found to be related with higher cancer risk in combination with BRCA1 or BRCA2 mutations.

The existence of genetic anticipation in Lynch syndrome (see section 1.3) is still under debate, as contradictory data has been published (reviewed in Bozzao et al. (2011b)); furthermore, significant results in the presence (or absence) of anticipation in LS highly depend on the statistical method used (reviewed in Boonstra et al. (2010)). Recently, Valls-Bautista et al. (2015) showed that in colorectal cancer patients, telomere lengths were altered in blood and normal mucosa, and Qin et al. (2014) found a significant association between short telomeres and colorectal cancer risk (especially in younger than 60 years patients). Conversely, long telomeres were related with cancer risk in familial colorectal type X patients that fulfill the Amsterdam criteria (see section 1.3.1) but do not show MMR deficiency or MSI (Seguí et al., 2014).

### 1.3 Colorectal cancer

Colorectal cancer (CRC) is the $4^{\text {th }}$ most common cancer in the UK and the $3^{\text {rd }}$ worldwide. A model for colorectal tumorigenesis was described by Fearon \& Vogelstein (1990) and more updated reviews ((Rustgi, 2007), (Pino \& Chung, 2010), (Kanthan et al., 2012) and (Centelles, 2012)) explain in detail the molecular mechanisms involved in colorectal carcinogenesis. Most of the CRC cases ( $95 \%$ ) are sporadic while only a $5 \%$ are inherited predispositions (Lynch \& de la Chapelle, 2003); of those sporadic CRC, the vast majority ( $60-80 \%$ ) arise as a MMR defect caused by somatic mutations (usually, promoter hypermethylation of MLH1). In addition, approximately $4 \%$ of the inherited CRC predisposition corresponds to Lynch syndrome (see details below) that is characterised by an inherited germline MMR mutation. Ahmed et al. (2013) genetically characterised 24 CRC cell lines describing the great heterogeneity in mutations and epigenetic regulation that can be found in this type of tumours (reviewed in Kim \& Kang (2014)). Furthemore, an extensive molecular description of the somatic mutations found in 224 CRC samples by next generation sequencing was published by Cancer Genome Atlas Network (2012) suggesting that the mutation rate might be a good prognostic signal as MSI tumours had a better survival. In addition to MSI, there are many other prognostic markers to study CRC (reviewed in Deschoolmeester et al. (2010)) and to identify the specific type.

### 1.3.1 Hereditary colon cancer

The hereditary predisposition for CRC is classified in: Lynch syndrome (LS) (previously known as hereditary non-polyposis colorectal cancer, HNPCC), accounting for $\approx 3$ to $4 \%$ of CRC and familial adenomatous polyposis (FAP) that accounts for nearly $1 \%$ of the hereditary CRC (reviewed in Lynch \& de la Chapelle (2003), Rustgi (2007) and Garre et al. (2014)).

Familial adenomatous polyposis cancer ( $\sharp 175100$ ), is an autosomal inherited disorder caused by mutations in the tumour suppressor gene
$A P C$ and is characterised by adenomatous polyps in the colon and rectum. Fodde et al. (2001a) suggested that mutations in the C-terminus of APC produce the chromosomal instability that characterises non-LS colorectal tumours. In addition, three autosomal recessive variants also account for this type of inherited CRC predisposition: MYH-associated polyposis (MAP) associated to MUTYH gene mutations ( $\ddagger 608456$ ), PeutzJeghers syndrome (PJS) associated to STK11 gene mutations ( $\ddagger 175200$ ) and juvenile polyposis syndrome (JPS) associated to BMPRIA or SMAD4 gene mutations ( $\ddagger 174900$ ) (reviewed in (Centelles, 2012).

Lynch syndrome ( $\sharp 120435$ ), is an autosomal dominant inherited disorder caused by a germline mutation in an MMR gene. It is characterized by an early-onset of CRC, often with a bias for the right colon and the presence of other extra colonic tumours such as urologic, upper gastrointestinal or gynaecologic cancer. The clinical hallmarks of this inherited disease resulted in a classification scheme, first designated Amsterdam Criteria I and II and later as Bethesda criteria and Revised Bethesda guidelines (reviewed in (Rustgi, 2007) and (Serrano et al., 2012)) that classify potential families for further screening. Genetic linkage studies identified two loci reporting $90 \%$ of HNPCC (Fishel \& Kolodner, 1995); one of them was described by Peltomäki et al. (1993) and mapped to 2p15-16 (corresponding to MSH2 gene) and accounts for $60 \%$ of LS; the other, was located in 3p21 (corresponding to MLH1 gene) and causing $30 \%$ of LS (for a review of MLH1 and MSH2 mutations refer to (Peltomäki \& Vasen, 1997)). Human tumour cell lines have been identified having MSH2 or MLH1 mutations and presenting mutator phenotypes and microsatellite instability (MSI) ((Bronner et al., 1994) and (Fishel \& Kolodner, 1995)). MLH6 and MLH3 germline mutations have also been found in LS patients although at lower frequencies than for MSH2 and MLH1 ((Kolodner et al., 1999) and (Wu et al., 2001)) and it was proposed that mutations in MSH6 predispose individuals to late onset familial colorectal carcinomas compared with the early onset familial colorectal carcinomas. Later, some MSH6 germline mutations were studied in pu-
tative HNPCC patients although the repair function was not affected by these mutations because they did not abolish MSH2-MSH6 interaction (Kariola et al., 2002). Very few mutations have been reported in other genes related to the MMR pathway such as EXO1 (Jagmohan-Changur et al., 2003) or PMS2 (reviewed in Peltomäki (2016)) suggesting that if related to LS, they might be involved in low penetrance cancer susceptibility. In addition, germline mutations in genes not directly related with the MMR pathway such as $P O L E$ (DNA polymerase $\epsilon$ ) have also been found to predispose to CRC (Elsayed et al., 2015). A genetic screen approach has been found useful for detecting the pathogenicity of mutations in MMR genes called VUS (variants of unclear significance), allowing to design diagnostic tools for LS patients (Drost et al., 2013). Nevertheless, inheriting a germline mutation in these genes or others related to the MMR pathway is not sufficient for tumour initiation, so heterozygosity in MMR genes is just a prerequisite for the development of cancer ((Hemminki et al., 1994) and (Chang et al., 2005)). The cytoplasmic location of the MSH2 protein in many HNPCC tumours, suggested that instead of an absence of the protein, the germline gene mutation produced an abnormal translocation of the protein to the nucleus (Fujiwara et al., 1998). Currently, it is accepted that LOH of tumour suppressor genes is one of the key steps to carcinogenesis in colorectal cancer (reviewed in Ozaslan \& Aytekin (2009)) and the loss of one allele at a constitutionally heterozygous locus indicates the probability of loosing a tumour suppressor gene, what might promote neoplastic progression. In addition, it has been found that $m i R-155$ overexpression can downregulate MLH1, MSH2 and MSH6 protein expression (reviewed in Yamamoto et al. (2012)) possibly contributing to LOH. Furthermore, telomerase activity in normal mucosa and lymphocytes in HNPCC patients has been found at higher levels compared to non-HNPCC (Cheng et al., 1998), suggesting that the MMR defect directly or indirectly might contribute to telomerase reactivation in normal tissues thus, increasing the cancer risk.

### 1.4 Hypothesis and main aims

Taking into account the preliminary data discussed above, in particular that MSH2 deficiency increases the telomere mutation frequency (measured as changes in the length of the telomere variant repeats), we hypothesised that defective mismatch repair could produce a trend towards losses of repeats from telomeres, so contributing to an increase in the telomere shortening rate. If $50 \%$ reduction in mismatch repair protein expression is sufficient to limit repair of tandem repeats during replication then, Lynch syndrome individuals carrying an MSH2 germline mutation, may show an increase in the telomere shortening rate compared with healthy individuals. This might be an explanation for the proposed genetic anticipation in this disease. To test our hypothesis, the main aims of this Thesis were:

1. To down-regulate the MSH2 protein in human primary cells grown in $5 \% \mathrm{O}_{2}$ and to compare the telomere shortening rates between clones with different MSH2 protein levels and control clones.
2. To analyse mutant telomeres in a colon cancer cell line with mutated MSH2 gene to determine whether MSH2 deficiency results in a bias towards loss of telomeric repeats, contributing to telomere shortening.
3. To test a telomerase inhibitor in the same colon cancer cell line, with mutated MSH2, to understand the impact of DNA mismatch repair deficiency on telomere length maintenance.
4. To compare telomere shortening with age in a cohort of healthy and Lynch syndrome individuals and to determine whether the inheritance of telomere length is affected in Lynch syndrome families.

## 2 Materials and methods

### 2.1 Materials

Antibodies: Mouse monoclonal (clone 3A2B8C) anti-MSH2 IgG (Abcam, UK); mouse monoclonal (clone 6C5) anti-GAPDH (Life Technologies, UK); sheep horseradish peroxidase (HRP) conjugated secondary anti-mouse (GE healthcare, UK).

Cells: Table 2.1 shows general information for the cell lines used in this Thesis.

|  | SW-480 | LoVo | WI38 |
| :--- | :--- | :--- | :--- |
| Organism | Homo sapiens | Homo sapiens | Homo sapiens |
| Tissue | Colon | Colon | Lung |
| Morphology | Epithelial | Epithelial | Fibroblast |
| Properties | Adherent | Adherent | Adherent |
| Donor age | 50 years | 56 years | 3 GM fetus |
| Ethnicity | Caucasian | Caucasian | Caucasian |
| Sex | Male | Male | Female |
| Disease | Duke's type B, colorectal | Duke's type C, grade IV, | Normal |
|  | adenocarcinoma | colorectal adenocarcinoma |  |
| MMR | proficient | msh2 - | proficient |
| TSG | apc- | - | - |
| Oncogenes | myc+, myb+, ras+, fos+, <br> p53+, sis+, abl-, ros-, src- | myc+, myb+, ras+, fos+, | none |
| Reference | (Witty et al., 1994) | (Umar et al., 1994) | (Hayflick \& Moor- |
|  |  |  | head, 1961) |

Table 2.1: Cell lines information. Data acquired and adjusted from the American Type Culture Collection (ATCC) website. "GM" gestation months; "TSG" tumour suppressor gene.

Cell staining: $\beta$-galactosidase staining kit (Cell signalling, UK); Propidium iodide (Sigma-Aldrich, UK); DAPI (Sigma-Aldrich, UK) and Trypan blue solution, $0.4 \%$ (Life Technologies, UK).

Chemicals and others: Deionized formamide (purity > 99.5\%) (Merck, UK); AccuGel (40\% Acrylamide:Bis-Acrylamide, 19:1) (Fisher Scientific, UK); ProtoGel (30\%) 37.5:1 acrylamide to bisacrylamide (Fisher Scientific, UK); Cloning discs (Sigma-Aldrich, UK); BLUeye prestained protein ladder (GeneFlow, UK); Hi-Di ${ }^{\text {TM }}$ formamide (Applied Biosystems, UK); MapMarker ${ }^{\circledR}$ X-Rhodamine labeled 1000 bp (ROX-1000XL) ladder (BioVentures, US); minimum essential medium (Thermo Scientific, UK); Dulbecco's modified Eagle medium (Thermo Scientific, UK); 100X non-essential amino acids (Thermo Scientific, UK); fetal bovine serum (Pan-Biotech, Germany); EDTA (Sigma-Aldrich, UK).

Drugs: Puromycin dihydrocloride (Sigma, UK); ampicillin (Sigma, UK) and BIBR1532 -telomerase inhibitor- (Damm et al., 2001) (Stratech Scientific, UK).

Enzymes: All enzymes were bought from New England Bio labs (NEB, UK).
Hybridization: MAGNA nylon transfer membrane, 0.45 microns (Fisher Scientific, US) for Southern and dot blots and Amersham Hybond ECL membrane (Fisher Scientific, UK) for western blots.

Kits: E.Z.N.A. ${ }^{\circledR}$ Plasmid Maxi Kit (Omega, UK); MaXtract High Density (Qiagen, UK); Zymoclean ${ }^{\text {TM }}$ Gel DNA Recovery Kit (ZymoResearch, UK); Oragene ${ }^{\text {TM }}$ DNA saliva kit (Genotek, US); Enhanced chemiluminescence western blotting detection system (ECL prime) (GE healthcare, UK); TRAPeze ${ }^{\circledR}$ telomerase tetection kit (Millipore, UK); Ingenio ${ }^{\circledR}$ electroporation kit for Lonza-Amaxa ${ }^{\circledR}$ Nucleofector ${ }^{\circledR} \mathrm{II} / 2 \mathrm{~b}$ devices (Mirus, UK); quick start ${ }^{\text {TM }}$ Bradford protein assay kit (BioRad, UK); Pellet Paint ${ }^{\circledR}$ (Millipore, UK); AmPure beads (Beckman Coulter, UK) and Dynabeads ${ }^{\circledR}$ His-Tag (Thermo Scientific, UK).

Oligonucleotides and plasmids: all primers were ordered from Sigma, UK, except for NED labeled TagTelX primer, that was ordered from Thermo Scientific, UK. pSuperior.puro and pSuperior.puro.shMSH2 plasmids were ordered from Oligoengine, US.

Saliva samples: Ethics approval was originally granted by Birmingham Women's Hospital and transferred to University Hospitals of Leicester NHS Trust for Lynch syndrome samples and granted by the University of Leicester for control samples. Written informed consent was obtained from all participants. Saliva samples were collected using the Oragene ${ }^{\text {TM }}$ DNA saliva kit (Genotek, US) that was sent by post to each participant together with all the documentation accepted by the Ethical committees.

- Control samples: Volunteers and their families were mostly identified from the Department of Genetics in the University of Leicester, UK. The recruitment process was set up by Carmen Garrido and Dr. Nicola Royle and samples were collected and processed by Carmen Garrido.
- Lynch syndrome (LS) samples: LS families were recruited from the Birmingham Women's Hospital (BWH) and from the Leicester Royal Infirmary (LRI). The recruitment process was set up by Dr. Frances Tippins (Univ. of Leicester iB.Sc. student 2013-2014) and Dr. Nicola Royle, in conjunction with Prof. Eamonn Maher (BWH) and Dr. Julian Barwell (LRI). Samples were collected by Dr. Frances Tippins (Univ. of Leicester iB.Sc. student 2013-2014) and Carmen Garrido (Univ. of Leicester, 2014-2016) with the collaboration of Jonathan Hoffmann and Wayne Glover (BWH). Only ten families were fully analysed by Dr. Frances Tippins (FT) and the remainder were processed and analysed by Carmen Garrido (CG), including the recruitment of 5 new BWH families. For a breakdown of the samples collected by FT or CG refer to appendix F.


### 2.2 Methods

### 2.2.1 Cell culture

Tissue culture was performed in a class II laminar flow hood in a designated tissue culture area. SW480 and LoVo adherent cells were grown in Dulbecco's modified Eagle medium (Thermo Scientific, UK) with $10 \%$ fetal calf serum (Sigma, UK). WI38 primary fibroblast cells were grown in modified Eagle medium (Thermo Scientific, UK) supplemented with 1X non-essential amino acids (Thermo Scientific, UK) and $20 \%$ fetal bovine serum (PAN Biotech, DE). Cells were grown in a $37^{\circ} \mathrm{C}$ incubator with ambient $\mathrm{O}_{2}(21 \%), 5 \% \mathrm{CO}_{2}$ and high humidity with the exception of the WI38 primary cells transfected with the shMSH2 construct that were grown at $37^{\circ} \mathrm{C}$ in a $5 \% \mathrm{O}_{2}$ and $5 \% \mathrm{CO}_{2}$ environment (Sanyo, UK). Subculture of adherent cells was performed when $80 \%-90 \%$ confluence was reached. Media was aspirated, and Trypsin-EDTA (Gibco, UK) added ( 1 ml or 2 ml for T25 or T75 cm flasks -Greiner bio-onerespectively). Fresh media was added when cells were completely detached. Cell counts or Trypan blue assays were made using a Neubauer hemocytometer as described below and cells were seeded in new flasks with fresh media using 1:2 to $1: 8$ splits depending on the cell line.

Cell pellets for DNA or protein extractions were snap-frozen at $-80^{\circ} \mathrm{C}$ after a five minutes centrifugation step at 420 x g, a PBS (Phosphate-Buffered Saline) wash and a five minutes spin at 420 xg (all tissue culture centrifugation steps were done using a Sorvall ${ }^{\mathrm{TM}}$ Legend RT centrifuge).

### 2.2.1.1 Cell counts and population doubling calculation

Cell counts were performed in a Neubauer hemocytometer as recommended by the manufacturer. Population doublings (PD) were calculated using the following equation as described in ATCC (2014):

$$
\begin{equation*}
\mathrm{PD}=\frac{\log 10 \text { Final Count }-\log 10 \text { Starting Count }}{0.301} \tag{2.1}
\end{equation*}
$$

### 2.2.1.2 Trypan blue exclusion

Trypan blue is a staining solution that penetrates the cellular membranes of dead cells, therefore it can be used to calculate percentage of viability in a cell culture. Cells resuspended in PBS were incubated with $1: 1$ (v/v) $0.4 \%$ Trypan blue for three minutes and cell counts were performed in a Neubauer hemocytometer. Cell viability was calculated using the following equation:

$$
\begin{equation*}
\% \text { Viability }=1-\frac{\text { Number of dead (blue) cells }}{\text { Total No. cells }} * 100 \tag{2.2}
\end{equation*}
$$

### 2.2.1.3 $\beta$-galactosidase

The $\beta$-galactosidase assay is widely used to assess the percentage of senescent cells in a cell culture. This assay indirectly detects the intrinsic $\beta$ galactosidase activity through a colorless substrate (X-gal) that turns blue when is degraded by this enzyme that is present and active at pH 4.0 in all cells. This assay uses a buffer at pH 6.0 at which none of the X-gal should become blue, but due to the higher proportion of $\beta$-galactosidase enzyme in senescent compared to non-senescent cells, some of the X-gal will be degraded by senescent cells producing a blue color. The protocol used was described in Debacq-Chainiaux et al. (2009). Cells ( $5 \times 10^{4}$ ) were seeded in 24 well plates and cultured for 24 or 48 hours before the assay. Cells were fixed with $2 \%$ formaldehyde and $0.02 \%$ glutaraldehyde in 1X PBS for three minutes and washed three times with 1X PBS. The staining solution ( 0.2 M citric acid/sodium phosphate $\mathrm{pH} 6.0,0.15 \mathrm{M} \mathrm{NaCl}, 2 \mathrm{mM} \mathrm{MgCl} 2,50 \mathrm{mM}$ potassium ferrocyanide, 50 mM potassium ferricyanide and $20 \mu \mathrm{~g}$ of X-gal in DMF -dimethylformamide-) was added to each well and the plate was incubated in a dry $37^{\circ} \mathrm{C}$ incubator for 15 to 16 hours. The staining solution was removed and three washes with 1X PBS were done. Fixation was done for 30 seconds with $100 \%$ methanol and air dry were followed by DAPI ( $1 \mu \mathrm{~g} / \mathrm{ml}$ in 1X PBS) staining for five minutes. DAPI was removed and fresh 1X PBS was added to
each well. Subsequently, the plate was analyzed by scanR software (version 2.4.0.11) to quantify the total number of cells (fluorescent DAPI staining) and senescent cells (brightfield blue-staining). The percentage of senescent cells was calculated as shown in the following equation:

$$
\begin{equation*}
\% \text { Senescence }=\frac{\text { Senescent cells }}{\text { Total No. cells }} * 100 \tag{2.3}
\end{equation*}
$$

### 2.2.1.4 Transfection

Plasmids for transfection were propagated in E. coli XL-blue (Mutant alleles are: recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1. Wild type alleles are: lac [F' proAB lacI $\left.{ }^{\mathrm{q}} \mathrm{Z} \Delta \mathrm{M}_{15} \operatorname{Tn} 10\left(\mathrm{Tet}^{\mathrm{r}}\right)\right]$. Bacteria were streaked on agar plates with $100 \mathrm{\mu g} / \mathrm{ml}$ ampicillin and incubated at $37^{\circ} \mathrm{C}$ in an air incubator over night. The following day, single colonies were grown in 5 ml of Luria broth (LB) media with $100 \mu \mathrm{~g} / \mathrm{ml}$ ampicillin at $37^{\circ} \mathrm{C}$ in a shaking incubator for 6 hours as a starter culture and 1 ml of the starter culture was inoculated into 100 ml of LB with $100 \mu \mathrm{~g} / \mathrm{ml}$ ampicillin that was grown at $37^{\circ} \mathrm{C}$ in a shaking incubator overnight. The next day, plasmid DNA was extracted by using the E.Z.N.A. ${ }^{\circledR}$ Plasmid Maxi Kit (Omega, UK). Plasmid quantification was assessed by Nanodrop (Thermo Scientific, UK) and plasmid quality was assessed by electrophoresis in a $0.8 \%$ LE agarose gel in 0.5 X TBE ( 89 mM Tris, 89 mM Boric Acid, 2 mM EDTA) with $0.5 \mu \mathrm{~g} / \mathrm{ml}$ ethidium bromide. Plasmids were linearized by digesting $20 \mu \mathrm{~g}$ of DNA with 10 units of BamHI at $37^{\circ} \mathrm{C}$ for a minimum of one hour. Linearization was assessed by electrophoresis in a $0.8 \%$ LE agarose gel as described before. Linear plasmids were precipitated by adding 0.25 mM sodium acetate at pH 5.6 and 2.5 volumes of $100 \%$ ethanol and eluted in pure water (SIGMA, UK).

Prior to transfection, WI38 cells were expanded to get the desired number of cells for parallel transfection reactions and conditioned media from each passage was collected, purified from debris by spinning five minutes at 420 xg and stored at $4^{\circ} \mathrm{C}$ until use. On the day of transfection, cells were counted and $1 \mathrm{x} 10^{6}$ cells were pelleted at 90 x g for ten minutes. Each cell pellet was mixed with $100 \mu \mathrm{l}$ of Ingenio ${ }^{\circledR}$ transfection solution (Mirus, UK) and $2 \mu \mathrm{~g}$ of plasmid
(either transient pmaxGFP for transfection efficiency or pSuperior.puro and pSuperior.puro-shMSH2 for stable clones). Electroporation was performed in 0.2 cm cuvettes in an Amaxa ${ }^{\circledR}$ Nucleofector ${ }^{\circledR}$ II/2b device using program V001 as it has been an efficient delivery method for other primary fibroblasts (Nakayama et al., 2007). After electroporation, cells were incubated in $500 \mu \mathrm{l}$ of pre-warmed media (1:1 ratio of normal:conditioned media) for ten minutes and transferred into $10 \mathrm{~cm}^{2}$ Petri dishes (Sarstedt, UK) adding 7.5 ml of media. For cells transfected with pmaxGFP plasmid for transfection efficiency, cells were diluted into 24 well plates with acid-treated \# 1 coverslips for subsequent cell screening in an Olympus fluorescence microscope. For the remaining cells transfected with pSuperior.puro or pSuperior.puro-shMSH2, three days after transfection, media was removed and selection with $0.75 \mu \mathrm{~g} / \mathrm{ml}$ of puromycin was added and media changed every three or four days. Clones were isolated using cloning discs (Sigma, UK) between one to two weeks after selection and transferred to 24 -well plates (Sarstedt, UK). Media was added every four days while checking for confluency. When $\approx 90 \%$ confluency was reached for a clone, it was transferred to 6 -well plate prior to taking a cell count as a starting PD and subsequent expansions were performed in T25 flasks. During approximately the next three months, sub-culturing and cell pellets for the dividing clones were performed alongside Trypan blue counts and $\beta$-galactosidase assays.

### 2.2.1.5 Telomerase inhibition

A stock solution of 10 mM of the telomerase inhibitor BIBR1532 (Stratech Scientific, UK) was diluted in $100 \%$ dimethyl sulfoxide (DMSO). For telomerase inhibition directly on crude cell extracts, a final concentration of $30 \mu \mathrm{M}$ was prepared using CHAPS buffer (3-[(3-cholamidopropyl)dimethylammonio]-1propanesulfonate) and the cell lysate was incubated on ice for fifteen minutes before preforming telomerase repeat amplification protocol (TRAP assay) as described in section 2.2.2.10. For telomerase inhibition on cell culture, a final concentration of 10 and $30 \mu \mathrm{M}$ were used (El-Daly et al., 2005) and (El Daly \& Martens, 2007). Aliquots of Dulbecco's modified Eagle medium (Thermo

Scientific, UK) with the indicated concentration of BIBR1532 were prepared once a week and the stock solution of drug was stored at $-80^{\circ} \mathrm{C}$. DMSO in an equivalent and non-cytotoxic concentration ( $0.3 \%$ ) to drug dilution was used as negative control. Every three-four days, cells were counted, harvested for FAC (Flow-cytometry analysis) and/or pelleted for DNA extraction and subcultured in fresh Dulbecco's modified Eagle medium (Thermo Scientific, UK) with 10 or $30 \mu \mathrm{M}$ of BIBR1532 or $0.3 \%$ DMSO.

### 2.2.1.6 Flow-cytometry analysis (FAC)

To perform FAC analysis, cells and supernatant were collected at each passage at a concentration of approximately $1 \times 10^{5}$ cells $/ \mathrm{ml}$, centrifuged 0.4 x g during five minutes at $4^{\circ} \mathrm{C}$, resuspended in $500 \mu \mathrm{l}$ of precooled $70 \%$ ethanol and stored at $-20^{\circ} \mathrm{C}$ until use. Cells were stained with $10 \mu \mathrm{~g} / \mathrm{ml}$ propidium iodide dissolved in 1X PBS and $5 \mu \mathrm{~g} / \mathrm{ml}$ of RNase for one hour at $37^{\circ} \mathrm{C}$. The FAC reading was conducted using a BD FACSCanto ${ }^{\text {TM }}$ II, the analysis was performed using the BD FACSDiva software and statistical analysis of the results and graphs were done using GraphPad Prism software, version 6.0a.

### 2.2.2 Molecular biology

### 2.2.2.1 DNA extraction from cell pellet

Cell pellets were resuspended in 1X filtered SSC ( 15 mM sodium citrate, 150 mM sodium chloride) buffer. For small ( $10^{4}$ to $10^{5}$ cells), $250 \mu \mathrm{l}$ of 1X SSC, and for bigger ( $>10^{6}$ cells) pellets, $500 \mu \mathrm{l}$ of 1X SSC were added and cell suspension split into two 2 mL eppendorf tubes. Cells were lysed by adding 250 $\mu \mathrm{l}$ of lysis solution ( 100 mM Tris- $\mathrm{HCl} \mathrm{pH} 7.5,100 \mathrm{mM} \mathrm{NaCl}, 10 \mathrm{mM}$ EDTA, $1 \%$ sarkosyl) and RNA was digested by adding $5 \mu \mathrm{l}$ of RNase ( $10 \mathrm{mg} / \mu \mathrm{l}$ ) at room temperature for twenty minutes. Proteins were digested by adding 100 $\mu \mathrm{g} / \mathrm{ml}$ of proteinase K and incubating at $55^{\circ} \mathrm{C}$ for 5 to 6 hours in a water bath. DNA extraction was made by adding $500 \mu \mathrm{l}$ of phenol:chloroform:isoamyl alcohol (25:24:1) and mixing. The organic and aqueous phases were separated by centrifugation at $15,600 \mathrm{x} \mathrm{g}$ for seven minutes at room temperature using

MaXtract High Density tubes (Qiagen, UK). The aqueous phase was transferred to a fresh eppendorf and DNA was precipitated using 0.2 M NaOAc ( pH 5.6 ) and 2.5 volumes of $100 \%$ ethanol (for very small pellets, $2 \mu$ of Pellet Paint ${ }^{\circledR}$ (Millipore, UK) was added to facilitate the visibility of the DNA pellet). The DNA pellet was transferred to a new tube and washed with $800 \mu \mathrm{l}$ of $80 \%$ ethanol. The pellet was dried at room temperature and DNA dissolved in the appropiate volume of pure water (SIGMA, UK). The DNA concentration was quantified using Nanodrop (Thermo Scientific, UK) and DNA quality was assessed by electrophoretic size separation in a $0.8 \% \mathrm{LE}$ agarose in 0.5 X TBE ( 89 mM Tris, 89 mM Boric Acid, 2 mM EDTA) with $0.5 \mu \mathrm{~g} / \mathrm{ml}$ ethidium bromide.

### 2.2.2.2 DNA extraction from saliva

DNA from Lynch syndrome patients and control families was extracted from saliva using the Oragene ${ }^{\mathrm{TM}}$ DNA saliva Kit (Genotek, US). Each sample was weighed after arrival and heated for 3 hours at $50^{\circ} \mathrm{C}$ in an air incubator to assure homogeneity. Only half of the sample ( $\sim 2 \mathrm{ml}$ ) was extracted and the other half was stored at $-80^{\circ} \mathrm{C}$ as a backup. A volume of $1 / 25^{\text {th }}$ of prepITL2P DNA extraction solution (Genotek, US) was added to the sample, the mix was incubated on ice for ten minutes and centrifuged for ten minutes at $2,500 \mathrm{x}$ g at room temperature in an Eppendorf 5804 centrifuge. Supernatant was transferred to a new tube and DNA was precipitated by mixing 2.5 volumes of $100 \%$ ethanol. After three minutes incubation at room temperature the sample was swirled gently to precipitate the DNA and incubated for ten additional minutes at room temperature. After a ten minutes centrifugation step at $2,500 \mathrm{x} \mathrm{g}$ at room temperature, supernatant was discarded and DNA pellet washed with $80 \%$ ethanol. Finally, the DNA pellet was dissolved in 500 $\mu \mathrm{l}$ of 5 mM PCR-clean Tris ( pH 7.5 ) and vortexed. DNA solution was incubated over night at room temperature to ensure complete DNA rehydration and concentration was estimated using a Nanodrop (Thermo Scientific, UK). Measurements for absorbances at 230, 260, 280 and 320 nm were recorded as well as 260/280 and 260/230 ratios. Corrected $\mathrm{A}_{260} / \mathrm{A}_{280}$ was calculated using
the following formula as recommended by the manufacturer:

$$
\begin{equation*}
\text { Corrected } \mathrm{A}_{260} / \mathrm{A}_{280}=\frac{\mathrm{A}_{260}-\mathrm{A}_{280}}{\mathrm{~A}_{260}-\mathrm{A}_{320}} \tag{2.4}
\end{equation*}
$$

High molecular DNA quality was assessed by size resolution in a $0.8 \% \mathrm{LE}$ agarose in 0.5X TBE with $0.5 \mu \mathrm{~g} / \mathrm{ml}$ ethidium bromide gel. DNA dilutions ( $10 \mathrm{ng} / \mathrm{\mu l}$ ) were prepared using 5 mM PCR-clean Tris ( pH 7.5 ) and stored at $-20^{\circ} \mathrm{C}$ for subsequent STELA analysis; prior to the analysis, DNA dilutions were anonimysed using letters in alphabetic order to eliminate any bias on the analysis related with the known age or pedigree location for each sample.

### 2.2.2.3 Sanger sequencing

For sequencing reactions, the Sanger method (Sanger et al., 1977) was used. PCR was performed to amplify the DNA fragment of interest and agarose gel electrophoresis was used to resolve the fragment by size. Gel extraction of the amplicon was performed using a Zymoclean ${ }^{\text {TM }}$ Gel DNA Recovery Kit (ZymoResearch, UK) following the manufacturer's protocol. For the sequencing reaction, $1 \mu \mathrm{l}$ of the Big Dye ${ }^{\mathrm{TM}}$ Terminator and $3.5 \mu \mathrm{l}$ of 5 x Big Dye ${ }^{\mathrm{TM}}$ Terminator buffer were mixed with the amplicon ( 20 ng of purified amplicon per kb of sequence) and $3.2 \mu \mathrm{M}$ of sequencing primer for a $20 \mu \mathrm{l}$ reaction volume. The reaction was cycled 28 times at $50^{\circ} \mathrm{C}$ annealing temperature for twenty seconds and $70^{\circ} \mathrm{C}$ extension time for two minutes. A cleaning step with filtered 2X SSC was followed by column purification and samples were sent to the protein nucleic acid chemistry laboratory in University of Leicester for sequencing. Traces analysis was performed using FinchTV.

### 2.2.2.4 DNA nanopore sequencing

1. DNA sample preparation: Amplicons to be sequenced were generated by PCR (for control sequences) or by STELA (for single telomeres), using the primers on table 2.2. After amplification, PCR products were purified using 1.2X of AmPure beads following the manufacturer protocol and eluted in pure water (SIGMA, UK) for a final concentration of $\approx$ $500 \mathrm{ng} / \mu \mathrm{l}$. The eluted product was measured by Nanodrop (Thermo Sci-
entific, UK) and resolved by size using $1 \%$ agarose gel electrophoresis to confirm concentration and DNA quality.

| Library | Primer pair |
| :--- | :--- |
| Tel1 | XpYpE2 + Telorette2 + Teltail |
| Tel2 | XpYp427G/415C + Telorette2 + Teltail |
| OL A31 | HHV6 probe 31 F + HHV6 probe 31 R |
| KK A49 | HHV6A probe 49 F + HHV6A probe 49 R |

Table 2.2: Amplicons and primers for nanopore sequencing
2. Pre-library: amplicons were prepared for sequencing using the Genomic DNA kit SQK-MAP004 from Oxford Nanopore as follows:
(a) End-repair using NEBNext ${ }^{\circledR}$ End Repair Module (NEB, UK) was performed for thirty minutes at $20^{\circ} \mathrm{C}$ in a Veriti ${ }^{\circledR}$ PCR device following the manufacturer's protocol for a $50 \mu \mathrm{l}$ reaction. Blunted DNA was again cleaned up using 1.2X of AmPure beads and eluted in $13 \mu \mathrm{l}$ of pure water (SIGMA, UK) in DNA LoBind eppendorfs.
(b) A-tailing using NEBNext ${ }^{\circledR}$ dA-Tailing Module (NEB, UK) was performed for thirty minutes at $37^{\circ} \mathrm{C}$ in a Veriti ${ }^{\circledR}$ PCR device following the manufacturer protocol for a $15 \mu \mathrm{l}$ reaction.
(c) Ligation of $15 \mu \mathrm{l}$ of $\mathrm{d}-\mathrm{A}$ tailed DNA with $5 \mu \mathrm{l}$ of adapter mix, $1 \mu \mathrm{l}$ of HP adapter (Oxford Nanopore, UK) and $25 \mu \mathrm{l}$ of Blunt/TA Ligase Master Mix (NEB, UK) was performed at room temperature for ten minutes. After ligation, the pre-library was cleaned up using 400 $\mu \mathrm{g}$ of Dynabeads ${ }^{\circledR}$ His-Tag (Thermo Scientific, UK) following the manufacturer's protocol and eluted in $13 \mu \mathrm{l}$ of pure water (SIGMA, UK) in protein LoBind eppendorfs.
3. Library: half of the pre-library ( $6 \mu \mathrm{l}$ ) was mixed with $140 \mu \mathrm{l}$ of EP buffer and $4 \mu \mathrm{l}$ of Fuel (Oxford Nanopore, UK) and loaded onto the minION ${ }^{\text {тм }}$ device (Oxford Nanopore, UK) after running platform QC (Quallity Control) for the device and washing twice with $150 \mu$ l of EP buffer. Analysis of the results was conducted using SPECTRE (Special Computational Teaching and Research Environment) as described in chapter 4.


Figure 2.1: Diagram of the library preparation steps for minION ${ }^{\text {TM }}$ sequencing. This diagram summarises the main steps involving library preparation. The sample preparation for genomic DNA (gDNA) consists on DNA fragmentation, but as we were only interested in one single telomere, we amplified it by STELA and all different telomere lengths behaved as fragmented gDNA.

### 2.2.2.5 Dot blotting

The dot blot was performed during the optimization of the telomere variant repeat (TVR) PCR to study primer labeling efficiency (see section 4.3.1). TS30T is a forward primer that, after being 5 ' end labeled with $\gamma^{32} \mathrm{P}-\mathrm{dATP}$, will amplify the A allele of the XpYp telomere together with variant repeat primers (see figure 2.3). Using different labeling conditions for the TS30T primer and blotting it against a template for the XpYp subtelomeric region, we predicted the best labeling conditions for the TVR. An amplicon template from the subtelomeric region of XpYp was prepared by PCR using primers XpYpB 2 and XpYpE 2 at $0.3 \mu \mathrm{M}$ final concentration and 10 ng of DNA extracted from the MSH2 ${ }^{-/-}$cancer cell line LoVo. The PCR was cycled 30 times as follows: 10 seconds at $96^{\circ} \mathrm{C}, 30$ seconds at $66.5^{\circ} \mathrm{C}$ and 50 seconds at $68^{\circ} \mathrm{C}$ with a final extension step of 10 minutes at $68^{\circ} \mathrm{C}$. The amplicon was resolved by size in $2.5 \%$ NuSieve agarose gel electrophoresis and a band of the expected size, 500 bp , was observed. The XpYp subtelomeric amplicon was diluted 1:2 and 1:4 in water and used as triplicates. Denaturing solution ( $0.5 \mathrm{M} \mathrm{NaOH}, 2 \mathrm{M} \mathrm{NaCl}$ and 25 mM EDTA) mixed with bromophenol blue was added to each amplicon ( 5 volumes per tube, $250 \mu \mathrm{l}$ ). After five minutes incubation at room temperature, the samples were loaded onto a nylon transfer membrane (Fisher, UK) using a dot blot apparatus Hybri-Dot manifold (Life Technologies, UK) previously assembled. Each individual well was washed with $150 \mu \mathrm{l}$ of $2 \mathrm{X} \mathrm{SSC}$. blot was dismantled, the membrane dried at $80^{\circ} \mathrm{C}$ for five minutes and UV cross-linked (UV $700 \mathrm{~kJ} / \mathrm{cm}^{2}$ ). The membranes were soaked in 3X SSC and introduced in the hybridisation bottles. The pre-hybridisation step was performed in the oven at $58^{\circ} \mathrm{C}$ for ten minutes with 2 ml of pre-warmed TMAC solution ( 3 M tetramethyl ammonium chloride, $0.6 \%$ sodium dodecyl sulfate -SDS-, 1 mM diNaEDTA, 10 mM sodium phosphate $\mathrm{pH} 6.8,5 \mathrm{X}$ Dehnardts solution and $4 \mu \mathrm{~g} / \mu \mathrm{l}$ yeast RNA). The TMAC solution was discarded and 2.5 ml of TMAC solution were added to each tube for the hybridisation step. To stop the labeling, $20 \mu \mathrm{l}$ of kinase stop solution ( 25 mM diNa EDTA, $0.1 \%$ SDS and $10 \mu \mathrm{M}$ ATP) were added to each probe prior to denaturation in a hot block at $100^{\circ} \mathrm{C}$ for five minutes. After denaturation, the probe (TS30T primer end la-
beled under different conditions) was added to the bottles to hybridise for one hour at $53^{\circ} \mathrm{C}$. The washing solution ( 3 M TMAC, $0.6 \% \mathrm{SDS}, 1 \mathrm{mM}$ diNaEDTA and 10 mM sodium phosphate pH 6.8 ) was prewarmed at $50^{\circ} \mathrm{C}$. Three 2.5 ml washes with the washing solution for ten minutes at $55^{\circ} \mathrm{C}$ and one extra of 3.5 ml washing solution were performed. Membranes were scanned using the Typhoon phosphoimager and quantified using ImageQuant software, version 7.0 (GE Healthcare, UK).

### 2.2.2.6 Western blotting

Cell lysates. Cell pellets (from $7 \times 10^{4}$ to $5 \times 10^{5}$ cells) were lysed by the addition of $25 \mu \mathrm{l}$ of lysis buffer ( 0.05 M HEPES $\mathrm{pH} 7.4,1 \%$ Triton-X-100, 0.1 M $\mathrm{NaCl}, 1 \mathrm{mM}$ PMSF -phenylmethylsulfonyl flouride- and 1X protease inhibitor cocktail) and incubated for twenty minutes on ice. Lysates were centrifuged at $15,600 \mathrm{x} \mathrm{g}$ for ten minutes to remove debris and supernatant was quantified using quick start Bradford protein assay (BioRad, UK). Lysates were diluted in PBS and Laemmli loading buffer ( $10 \%$ glycerol, $0.02 \%$ bromophenol blue, $0.05 \% \beta$-mercaptoethanol, $2 \%$ SDS, 60 mM Tris pH 6.8 ) at a final concentration of $1 \mu \mathrm{~g} / \mu \mathrm{l}$ of protein. Samples were boiled at $95^{\circ} \mathrm{C}$ for five minutes and quickly spun prior to gel loading. Lysates were aliquoted and stored at $-20^{\circ} \mathrm{C}$ for subsequent use.

SDS-PAGE. Stacking gel ( $4 \%$ acrylamide, $0.1 \%$ bis-acrylamide, $0.1 \%$ SDS and 1.6 M Tris, pH 6.8 ) and running gel ( $10 \%$ acrylamide, $0.3 \%$ bis-acrylamide, $0.1 \%$ SDS and 2 M Tris, pH 8.8 ) were polymerised by the addition of APS -ammonium persulfate- ( $0.3 \%$ or $0.8 \%$ for running and stacking gel respectively) and TEMED ( $0.08 \%$ final concentration). Each sample ( 10 ng ) was loaded into each well and $2 \mu \mathrm{l}$ of BLUeye Prestained Protein Ladder were used as a size marker. The gel was run in an electrophoresis chamber (BioRad, UK) at 100 V for stacking gel and 150 V for running gel (until the bromophenol blue dye reached the bottom of the gel).

Western blot preparation. The membrane and the gel were incubated for ten minutes in transfer buffer ( 25 mM Tris, 192 mM glycine, $20 \%$ ethanol) and the protein transfer was carried out in semi-dry conditions using the

Trans-Blot ${ }^{\circledR}$ BioRad, UK) onto a nitrocellulose membrane (Amersham-ECL), (GE Healthcare, UK) at 1 mAmp per $\mathrm{cm}^{2}$ for one hour. Ponceau staining solution $0.1 \%$ (Sigma, UK) was used to check the transfer efficiency, the homogeneity of protein concentration and the evenness of loading between tracks.

Protein detection. The blot was cut in two small blots, each containing the size range for MSH2 ( 98 KDa ) or GAPDH ( 20 KDa ) proteins. As blocking agent, $5 \%$ milk powder (Marvel), dissolved in PBST (Phosphate-Buffered Saline with $0.1 \%$ Tween 20) was used for 45 minutes. Blots were then incubated with the respective primary antibody dissolved in $5 \%$ milk in PBST at specific dilutions (1:1,000 for MSH2 and 1:10,000 for GAPDH) for one hour at room temperature. Following binding of the primary antibody, the membrane was washed three times in PBST, for ten minutes each. Incubation with the secondary anti-mouse (to detect MSH2 and GAPDH) antibody conjugated to horseradish peroxidise (GE healthcare, UK) in a 1:10,000 dilution was for 45 minutes. The membrane was washed with PBST as before. The blot was developed using ECL prime western blotting detection system (GE healthcare, UK) during five minutes, the blot dried with 3 mm Whatman filter paper and covered with plastic wrap. The signal was visualized by exposing the membrane to an X-ray film in a cassette. Quantitative analysis of the resulting bands was performed using the ImageJ software, version 1.46.

### 2.2.2.7 Southern blotting

The $0.8 \%$ LE agarose gels were trimmed to the desired size with a scalpel and washed, shaking in depurinating ( 0.25 M HCl$)$ solution for seven minutes, rinsed in distilled water, incubated twice in denaturing ( 0.5 M NaOH , 1 M NaCl ) solution for ten minutes and twice in neutralising ( 0.5 M Tris, 3 $\mathrm{M} \mathrm{NaCl})$ solution for ten minutes. MAGNA nylon transfer membrane, 0.45 microns was incubated in 6X SSC for ten minutes before setting up the blot. The blotting apparatus was set up with 20X SSC in a tray with a glass plate and a piece of 3 mm Whatman paper. After 4 to 5 hours, the apparatus was dismantled and the membrane dried at $80^{\circ} \mathrm{C}$ in an air incubator. DNA was UV crosslinked at $700 \mathrm{~kJ} / \mathrm{cm}^{2}$ in a UVP UV crosslinker.

### 2.2.2.8 Single telomere length analysis (STELA)

This technique was used to amplify individual telomeres and to measure their length using a modification of the original protocol in (Baird et al., 2003). Extracted DNA as show in sections 2.2.2.1 or 2.2.2.2 was diluted at a final concentration of $250 \mathrm{pg} / \mu \mathrm{l}$ in a solution containing $225 \mathrm{ng} / \mu \mathrm{l}$ of Telorette 2 and $0.75 \mathrm{ng} / \mu \mathrm{l}$ of carrier yeast tRNA (for primer sequences refer to appendix). Telorette 2 consists of 20 nucleotides, which complement the 3 ' overhang of the telomere (and share the same sequence as Teltail), and 7 nucleotides of unique sequence. Long-range PCR was carried out using Teltail primer and a chromosome-specific primer located in the subtelomeric region. The cycling conditions were as follow: 24 cycles of $96^{\circ} \mathrm{C}$ for 10 seconds, with primerspecific annealing temperatures for 30 seconds and extension at $68^{\circ} \mathrm{C}$ for 12 minutes. DNA fragments were resolved in $0.8 \%$ LE agarose in 0.5 X TBE, 0.5 $\mu \mathrm{g} / \mathrm{ml}$ ethidium bromide gel and Southern blotted as described before. Prehybridisation and hybridisation was carried out in modified Church buffer ( 0.5 M sodium phosphate buffer, 1 mM EDTA, $7 \% \mathrm{SDS}$ ) at $65^{\circ} \mathrm{C}$ and a randomprimed $\alpha^{32} \mathrm{P}-\mathrm{dCTP}-l a b e l e d$ telomere and ladder ( 1 kb and HR ) probes added to the hybridisation. Membranes were washed in high stringency buffer ( 0.5 X SSC, $0.1 \%$ SDS) and hybridized fragments were detected using a Typhoon 9400 PhosphorImager (GE Healthcare, UK). The results were analysed using ImageQuant software, version 7.0 (GE Healthcare, UK) for sizing and GraphPad Prism software, version 6.0a for statistical analyses.


Figure 2.2: Diagram of single telomere length analysis (STELA). Adapted from Baird et al. (2003).

### 2.2.2.9 Telomere variant repeat PCR (TVR)

This method described in (Baird et al., 1995) allows determination of the interspersion of the consensus telomere repeat (TTAGGG) with sequence variant repeats (such as TCAGGG or TGAGGG among others), present at the start of some telomere alleles. Here, single molecule STELA products (smSTELA) were generated to identify telomere molecules with mutations in those repeat variants. Genomic DNA ( 70 to 150 pg per reaction) was amplified by STELA (as described before 2.2.2.8) in $20 \mu \mathrm{l}$ PCRs. PCR products ( 5 to $6 \mu \mathrm{l}$ ) were resolved by size using $0.8 \%$ LE agarose gel electrophoresis, Southern blotted and hybridized to the telomeric probe to assess smSTELA that would be subsequently analysed by TVR.

1. Denaturing acrylamide gels: in the original protocol (Baird et al., 1995), $0.1 \mu \mathrm{M}$ of forward primer was 5 ' end labeled with $\gamma^{32} \mathrm{P}$-dATP and the PCR performed with $0.4 \mu \mathrm{M}$ of each variant repeat primer. PCRs reactions were cycled 19 times for 40 seconds for the annealing step and 2 minutes for the extension step at $70^{\circ} \mathrm{C}$. After the PCR, formamide loading dye ( $95 \%$ formamide, 10 mM EDTA pH 8.0, $0.1 \%$ bromophenol blue and $0.1 \%$ xylene cyanol) was added to each sample and a five minute denaturation step at $96^{\circ} \mathrm{C}$ was performed in a thermocycler. The PCR products were resolved in denaturing acrylamide gels ( $6 \%$ acrylamide, 7.6 M urea in 1X TBE) polymerised with TEMED and APS. Gels were pre-run using 1X TBE as running buffer at constant 60 W for twenty minutes before loading the samples. When the gel electrophoresis was finished (different times depending on telomere and number of repeats), acrylamide gels were stuck to two pieces of 3 mm Whatman paper and were dried at $80^{\circ} \mathrm{C}$ in a gel drier (BioRad, UK) for two hours. The image of the radioactive TVR gels was detected using a Typhoon 9400 PhosphorImager (GE Healthcare, UK) and manually analysed.
2. Capillary electrophoresis: the original protocol has several limitations (see table 2.3) that we tried to overcome modifying it to make it suitable for capillary electrophoresis. Using the same primer sequences,
the variant repeat detector primers were fluorescently labeled. Different fluorophores were used for each variant repeat primer, so allowing all TVR products to be separated in the same capillary gel if required.

Primer concentrations were as in the original protocol ( $0.4 \mu \mathrm{M}$ for each reverse primer and $0.1 \mu \mathrm{M}$ of forward primer) and PCRs were cycled 24 times for 40 seconds for the annealing step and 2 minutes for the extension step at $70^{\circ} \mathrm{C}$. After the PCR, $1 \mu \mathrm{l}$ of PCR product was added to $9 \mu \mathrm{l}$ of $\mathrm{Hi}-\mathrm{Di}^{\text {TM }}$ formamide (Applied Biosystems, UK) with ROX-1000XL ladder (BioVentures, US) into a 96 -well plate. PCR products were denatured for five minutes at $96^{\circ} \mathrm{C}$ in a termocycler and run into the ABI 3100 Genetic Analyzer. Settings and conditions for the run are shown in table 2.4 and figure 2.3 graphically compares the two methods. Results were analysed using GeneMapper v4.1 (Applied Biosystems, UK).

|  | Sequencing gels | ABI 3130 |
| :--- | :--- | :--- |
| Samples per run | $<50$ | up to 192 |
| PCR volume per run | $>5 \mu \mathrm{l}$ | $1 \mu \mathrm{l}$ |
| Time per experiment | 2 to 3 days | 4 to 5 hours |
| Hazards | Formamide <br> Radiation <br> Leaking | Formamide |
| Voltage | None | Size ladder |
| Sizing | Manual | Gene Mapper software |
| Analysis | $<2$ weeks | several months at $-20^{\circ} \mathrm{C}$ |
| Labeling stability | Radiation activity <br> Probe labeling efficiency <br> Gel mixture <br> Gel electrophoresis | Polymer <br> Ladder |
| Variable factors |  |  |

Table 2.3: TVR methods comparison. The table compares the performance of TVR using sequencing gels or the ABI 3130 Genetic Analyzer.

| Condition | Setting |
| :--- | :--- |
| Injection voltage | 1.2 kVolts |
| Run voltage | 15 kVolts |
| Pre run time | 180 sec |
| Injection time | 40 sec |
| Voltage step interval | 15 sec |
| Run time | 1200 sec |
| Polymer | POP- ${ }^{\mathrm{TM}}$ (Thermo Scientific, UK) |

Table 2.4: Conditions for the ABI 3100 genetic analyzer.

1. TVR-PCR (Radiation method)

Figure 2.3: Diagram of TVR. This diagram summarises the main steps of the telomere variant repeat (TVR) protocol, comparing the method using $5^{\prime}$ end labeled with $\gamma^{32} \mathrm{P}$-dATP on the left and fluorescently 5 ' end labeled primers on the right.

### 2.2.2.10 Telomerase repeat amplification protocol (TRAP)

Adaptation of the original protocols in Kim et al. (1994) and Piatyszek et al. (1995) was made to measure telomerase activity. Cell pellets stored at $-80^{\circ} \mathrm{C}$ were lysed with 1X CHAPS lysis buffer for twenty minutes on ice, centrifuged at $12,000 \mathrm{x} \mathrm{g}$ for twenty minutes and supernatant analysed by quick start Bradford protein assay (BioRad, UK) to measure protein concentration (Bradford, 1976). The TS primer was end-labeled in the presence of $\gamma-{ }^{32} \mathrm{P}$-ATP and T 4 polynucleotide kinase ( 0.5 units/ $\mu \mathrm{l}$ ) for twenty minutes. The T4 polynucleotide kinase was inactivated by five minutes heat shock at $95^{\circ} \mathrm{C}$. The reaction was performed with a final protein concentration of 100 to $500 \mu \mathrm{~g} / \mathrm{\mu l}$ using TRAPeze ${ }^{\circledR}$ telomerase detection kit (Millipore, UK). Heat inactivated control lysates were prepared by incubation in a water bath at $95^{\circ} \mathrm{C}$ for ten minutes. The extension of the TS primer by the telomerase present in the cell lysates was at $30^{\circ} \mathrm{C}$ for thirty minutes and the PCR cycling was: 28 cycles of $94^{\circ} \mathrm{C}$ during 30 seconds and $55^{\circ} \mathrm{C}$ during 33 seconds for amplification (see figure 2.4 for further information). DNA fragments were resolved in $6 \%$ denaturing ( 7.7 M of urea) acrylamide gel. After drying the gel for two hours at $80^{\circ} \mathrm{C}$ in a gel dryer (BioRad, UK), the radioactive bands were detected using a Typhoon 9400 PhosphorImager (GE Healthcare) and the image visualized using ImageQuant version 7.0 (GE Healthcare, UK).

### 2.2.3 Statistical analyses

All the statistical tests were performed using GraphPad Prism software, version 6.0a and a p value below 0.5 was considered significant. For each purpose, distinct tests were used; briefly:

1. Linear regression: was used to compare slopes between two or more groups, producing an ANCOVA (analysis of covariance) $p$ value for the slopes and the Y intercepts (if slopes are not different). In chapter 3 it was used to compare cell growth at 5 and $21 \% \mathrm{O}_{2}$, rate of cell division in control and shMSH2 clones and the relationship between telomere shortening rate and initial telomere length. In chapter 5 it was used to
compare cell growth curves between treated and untreated cancer cells. In chapter 6 it was used to establish the telomere shortening rate with age in control and Lynch syndrome individuals and to compare between the different groups.
2. Non-linear regression: was used in chapter 3 to draw a best-fit curve in those XY graphs that did not adjust to a linear curve.
3. Groups comparison: the non-parametric two tailed Mann-Whitney t-test was used to compare two groups while one-way ANOVA was used for more than two groups. When the comparison focused on the distribution (rank) of the data, non-parametric Kruskal-Wallis ranking test was performed, for example for comparisons of senescence and viability between control and shMSH2 clones (chapter 3), for comparisons of the age distribution of the saliva samples among the groups (chapter 6) and before pooling data from replicates for STELA. Finally, when more than two groups wanted to be compared, but each comparison stood alone, the ANOVA was uncorrected for multiple comparisons and Fisher test was used.

## 1. Elongation step (30 mins)


2. Telomerase Repeat Amplification Protocol (30 mins)


Figure 2.4: Diagram of TRAP assay. This diagram summarises the main steps in the telomerase repeat amplification protocol (TRAP). As the reverse primer 1 can anneal in any repeat, the PCR products would be shown as a 6 bp ladder as described by Kim et al. (1994).

## 3 Telomere dynamics in a human primary cell under MSH2 deficiency

### 3.1 Background

Telomeres of human fibroblast cells grown in culture shorten with each cell division between 30-200 base pairs (bp) ((Harley et al., 1990) and reviewed in Verdun \& Karlseder (2007)). Therefore, a negative relationship between telomere length and replicative capacity of human primary fibroblasts was predicted (Allsopp et al., 1992). The maximum number of cell divisions that a primary cell line can undergo before entering replicative senescence, the natural state of cell proliferation arrest, is known as the "Hayflick limit" (Hayflick \& Moorhead, 1961) and telomere length, specially short telomeres, and not the number of cell divisions, regulate the onset of telomere driven replicative senescence ((Sitte et al., 1998) and (Munro et al., 2001)). On the contrary, cancer cells maintain their telomere length by activating either the telomerase enzyme (Blackburn et al., 1989) or the Alternative Lengthening of Telomeres (ALT) pathway (reviewed in de Lange (2009) and Takubo et al. (2010)).

Telomeres are very susceptible to oxidative damage because of their high guanine content, being 8-oxo-7,8-dihydroguanine (8-oxoG) and 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) the two main DNA lesions produced in the presence of reactive oxygen species (ROS) (Wang et al., 2010). During replication, 8 -oxoG might bind either to cytosine or adenine; in the first case, the lesion remains silent but in the second case, a G:C to A:T transversion might occur (Bridge et al., 2014). In addition, oxidative stress increases the frequency of single-stranded breaks (SSB) throughout the genome but preferentially on telomeres (Petersen et al., 1998) and it has been shown that the SSB repair on telomeres is not as efficient as in other regions such as microsatellites (Coluzzi et al., 2014). Previously, it was demonstrated that hyperoxic $\left(40 \% \mathrm{O}_{2}\right)$ conditions increased telomere attrition rate (from 90 to 500 bp per
population doubling) and impeded cell proliferation after a few cell divisions (von Zglinicki et al., 1995). Besides, it was demonstrated that low oxygen conditions (Balin et al., 2010) and/ or high antioxidant levels (Lorenz et al., 2001) delayed the onset of replicative senescence and reduced the TSR (Serra et al., 2003) possibly due to the decline in the number of SSBs (von Zglinicki et al., 2000a).

Oxidative stress is related with both, replicative (RS) and induced senescence (IS) (see section 1.2.2.2). On the one hand, elevated ROS concentrations increase telomere shortening as demonstrated by von Zglinicki et al. (1995), blocking cell proliferation after few cell divisions. On the other hand, the use of antioxidants can revert senescence phenotypes such as endothelial dysfunction (Bhayadia et al., 2016), what is only possible in reversible IS but not RS. Interestingly, it was demonstrated by Britt-Compton et al. (2009b) that under low ( $3 \%$ ), but not normobaric ( $21 \%$ ) oxygen conditions, senescence was triggered by telomere shortening (overall mean of $83 \mathrm{bp} / \mathrm{PD} \pm 23 \mathrm{SD}$ ) in IMR90 fibroblasts while in MRC5 fibroblasts, senescence was triggered by telomere attrition at both oxygen conditions. This suggested that different cell lines might have distinct responses under oxygen stress as it was shown in ex vivo cultures of human CD34 ${ }^{+}$cells (Fan et al., 2008) and more recently by transcriptional assays in two human cancer cell lines (the hepatocellular carcinoma, Hep G2 and the colorectal adenocarcinoma, Caco2) (Deferme et al., 2015). In addition, primary fibroblasts (MRC5) growth-arrested by confluency exhibited a higher telomere shortening rate (TSR) when the proliferation block was released and accumulated SSB on the telomeres (Sitte et al., 1998). Therefore, it was suggested that accumulation of SSB on telomeres was the main trigger for the increased TSR observed under hyperoxia ((von Zglinicki et al., 2000a) and (Honda et al., 2001)).

The DNA MMR pathway repairs single base mismatches and insertion/deletion loops as well oxidative lesions through interaction of the MSH2-MSH6 complex with monoubiquitinated PCNA and DNA polymerase $\eta$ (Zlatanou et al., 2011) and it has been described that colorectal cancer cells deficient for the MSH2 protein have a higher telomeric mutation frequency in the vari-
ant repeat region of the telomere (Pickett et al., 2004). Interestingly, a synthetic lethality has been described between MMR gene mutations and DNA polymerases: $\mathrm{MSH}^{-/-}$combined with $\mathrm{POLB}^{-/-}$or MLH1 ${ }^{-/-}$combined with POLG ${ }^{-/-}$leading to the accumulation of oxidative lesions (Martin et al., 2010).

To understand the effect that MSH2 depletion had on the telomeres of a primary cell line, Mendez-Bermudez \& Royle (2011) firstly transfected the human primary lung fibroblast cell line (CCD34-Lu) with the catalytic subunit of telomerase (hTERT) (as it is known that telomerase can extend the life span of primary fibroblasts in culture by lengthening their telomeres (Bodnar et al., 1998) specially the shortest ones (Ouellette et al. (2000) and Britt-Compton et al. (2009a)). Unexpectedly, despite detectable telomerase activity, telomere length continued to shorten in the CCD34-Lu-hTERT cells cultured at $21 \%$ $\mathrm{O}_{2}$. Secondly, the MSH2 protein was depleted using an shRNA and it was found that telomeres in CCD34-Lu-hTERT clones with depleted MSH2 shortened at a significantly higher rate $242 \pm 13$ bp per Population Doubling (PD) compared to control clones ( $51 \pm 6 \mathrm{bp} / \mathrm{PD}$ ) suggesting that the MMR pathway was involved in the telomere length maintenance during replication.

Nevertheless, there were some unanswered questions arising from that experiment: a) was the effect cell line specific or do other primary cell lines show increased telomere shortening rate (TSR) in upon depletion of MSH2?, b) why were telomeres shortening even when telomerase was expressed?, c) was the TSR affected by the presence of telomerase?, and d) was the increased TSR in MSH2 depleted clones caused by the oxidative stress imposed by growing the cells at normobaric $\left(21 \% \mathrm{O}_{2}\right)$ conditions?

### 3.2 Aims

To answer the previous questions, stable clones expressing the same shMSH2 construct described in Mendez-Bermudez \& Royle (2011) were generated but using a different telomerase negative human primary lung fibroblast cell line (WI38) that is known to undergo many cell divisions before entering senescence. This eliminated the need to transfect the WI38 cells with hTERT prior to depletion of MSH2. To dismiss oxidative stress as a contributory factor, the WI38 cell line and clones generated after transfection were grown under 5\% $\mathrm{O}_{2}$ conditions. The objectives for this aim were:

- To transfect WI38 cells with a plasmid expressing shMSH2 to generate stable clones.
- To grow independent MSH2-depleted stable clones and to follow their growth dynamics.
- To measure the level of MSH2 expression by western blot and to select the clones with the highest MSH2 downregulation for further analysis.
- To measure telomere shortening rates (TSR) using single telomere length analysis (STELA) on clones with reduced MSH2 expression and control clones.
- To compare data with that obtained in a different cell line (MendezBermudez \& Royle, 2011) to overcome the limitations from the previous experiment.


### 3.3 Results

### 3.3.1 Experiment design

Normoxic (5\%) oxygen conditions were used to confirm that the higher TSR described in (Mendez-Bermudez \& Royle, 2011), when the cells were grown under $21 \% \mathrm{O}_{2}$ conditions, was solely due to depletion of the MSH2 protein and not to the high levels of oxidative stress. As previously described (von Zglinicki et al., 1995), the untransfected WI38 cell line grew faster and underwent a larger number of population doublings (PD) at $5 \% \mathrm{O}_{2}$ compared to $21 \% \mathrm{O}_{2}$ conditions (figure 3.1). Interestingly, growth curves comparisons between both oxygen conditions were not statistically significant ( $p=0.05$ ) for "younger" cells ( $\mathrm{PD}=25$, figure 3.1.A) but were highly significant ( $\mathrm{p}<0.0001$ ) for "older" cells $(\mathrm{PD}=32$, figure 3.1.B), confirming that low oxygen conditions delay the onset of senescence in WI38 cells (Balin et al., 2010). In addition, the slopes for the growth curves at $\mathrm{PD}=25$ and $\mathrm{PD}=32$ grown under the same oxygen conditions were statistically ( $p<0.0001$ ) different for both, $5 \%$ and $21 \% \mathrm{O}_{2}$ conditions suggesting that "older" primary cells divided more slowly than "younger" ones despite the oxygen concentration (due to the higher proportion of senescent cells in the former). Those results were consistent with a) experiments performed by Betts et al. (2008) showing that senescence can be delayed under low oxygen conditions even when telomeres are very short and b) the positive relationship between production of antioxidant enzymes such as glutathione (GST) and oxygen tension that occurs in log-phase ("younger") cells but is reduced in senescent ("older") cells (Balin et al., 2010).


Figure 3.1: WI38 growth curves under different oxygen conditions. A shows the growth curve for a WI38 "young" passage ( P 25). B shows the growth curve for a WI38 "old" passage (PD 32). For both graphs, blue colour refers to normoxic (biological) oxygen conditions ( $5 \% \mathrm{O}_{2}$ ) and green colour refers to normobaric (ambient) oxygen conditions ( $21 \% \mathrm{O}_{2}$ ). Best fit curves for each oxygen condition are shown on grey colour. "PD" = Population Doubling.

Figure 3.2 shows dose-response curves for WI38 cells grown in two different puromycin concentrations ( $0.5 \mu \mathrm{~g} / \mathrm{ml}$ and $1 \mu \mathrm{~g} / \mathrm{ml}$ ) and untreated cells as a control, to asses the best conditions for puromycin selection after transfection with the pSuperior.puro and pSuperior.puro.shMSH2 plasmids. The optimal concentration would be the lowest that kills most of the cells after one day. Following this assessment, colony selection was conducted with puromycin at $0.75 \mu \mathrm{~g} / \mathrm{ml}$ in subsequent experiments.


Figure 3.2: Puromycin dose-response curve for WI38 cells. The control curve shown in purple represents WI38 cells grown in media (MEM $+20 \%$ FBS +1 X non essential amino acids) without antibiotic. Red and green curves represent WI38 cells treated with $0.5 \mu \mathrm{~g} / \mathrm{ml}$ and $1 \mu \mathrm{~g} / \mathrm{ml}$ of puromycin respectively.

The plasmids used for transfection (pSuperior.puro and pSuperior.puroshMSH2), taken from Mendez-Bermudez \& Royle (2011) are shown in figure 3.3.A. As previously described in section 2.2.1.4, the plasmids were propagated in bacteria ( $E$. coli XL-blue) and extracted using maxi prep extraction
kits. The sequence across the shMSH2 insert was confirmed by Sanger sequencing as shown in figure 3.3.B. The transfection was performed in an Amaxa ${ }^{\circledR}$ Nucleofector ${ }^{\circledR} \mathrm{II} / 2 \mathrm{~b}$ device as described in section 2.2.1.4. Puromycin selection was added three days after transfection and the delivery efficiency was measured as the total number of colonies formed per million of transfected cells (figure 3.4.A). Two transfection experiments were performed, the first of which was not successful in terms of surviving clones due to contamination but the number of colonies was scored for subsequent comparisons. In figure 3.4.A, blue bars represent the first, unsuccessful experiment and green bars the second. The significant increase in the number of colonies was obtained by diluting each transfection reaction into three 10 cm Petri dishes (second experiment) instead of one. For transfections with the shMSH2 plasmid, an increase from 20 to 134 colonies per million of transfected cells was obtained and for the control plasmid, the increase was from 19 to 151 colonies per million cells. In addition, co-transfection with pmaxGFP plasmid with either pSuperior.puro control or pSuperior.puro-shMSH2 plasmids was used to estimate delivery efficiency by counting the percentage of green cells in reference to the total cell number (figure 3.4.C).


Figure 3.3: Plasmid and shRNA used for transfection. A pSuperior.puro-shMSH2 plasmid map. It contains an ampicillin resistance (AmpR) gene that will be expressed in E. coli cells for plasmid amplification and a puromycin resistance (PuroR) gene that will be expressed in mammalian cells for selection of transfected cells. The control plasmid (pSuperior.puro) does not contain the shMHS2 sequence. B shows the location of the shMSH2 under the H1 promoter was confirmed by Sanger sequencing using pSup1primer (see table in appendix B). The sequence chromatogram of 60 nucleotides corresponding to the shMSH2 is shown. $\mathbf{C}$ is the structure of the shMSH2, showing the complementarity between 19 nucleotides (small letters), the loop (capital letters) and both, 5 ' and 3 ' tails (capital letters).


Figure 3.4: Transfection efficiency. A is a bar graph showing the transfection efficiency measured as number of colonies per million transfected cells between two independent experiments (blue and green bars) for both, shMSH2 $\left({ }^{* * *} \mathrm{p}=0.0005\right)$ and control $\left({ }^{* *} \mathrm{p}=0.0091\right)$ plasmid. Ranks were compared by Mann Whitney test. B is an example of the green fluorescence signal produced by the co-transfection of shMSH2 and pmaxGFP plasmids in one positive cell compared with two non-positive cells which nuclei were stained with $1 \mu \mathrm{~g} / \mathrm{mL}$ DAPI. C shows quantification of the transfection efficiency as a measurement of green-positive cells in relation to green-negative cells expressed as a percentage. Green bars represent values for transfection with only pmaxGFP plasmid; purple, for pSuperior.puro-shMSH2 co-transfected with pmaxGFP plasmid and blue for pSuperior.puro control vector co-transfected with pmaxGFP plasmid. Table below the graph shows the number of cells with green signal over the total number of scored cells per day. Day 1 and 3 represent one and three days after transfection respectively. Selection was added on day 3 after scoring cells for that day. Day 4,6 and 11 represent one, three and seven days after puromycin selection respectively.

### 3.3.2 WI38 MSH2 $^{+/-}$clones growth

Due to the elevated mutation rate reported in MMR-deficient cells across the genome and in particular at microsatellite regions (Coolbaugh-Murphy et al., 2010) and at telomeres (Pickett et al., 2004), it was anticipated that fewer shMSH2 clones would survive or that they may grow more slowly and so more of these clones were isolated compared to control clones. In total, 33 shMSH2 clones were isolated using cloning discs, 23 of which grew successfully, giving a recovery efficiency of $70 \%$. For the control clones, 18 were isolated, 7 of which were subsequently followed up for cell growth. In addition, 10 shMSH2 and 5 control populations were grown in parallel to clones to study the effect that clonal variation had on growth curves. To draw growth curves, Population Doublings (PDs) were calculated following the equation 2.1.

Each clone was considered to be derived from a single cell (as sufficient dilution after transfection should allow colony formation from a single cell progenitor) and a $100 \%$ recovery was assumed following trypsinization using cloning discs to isolate independent clones. The first cell count was made at the trypsinization step in an $80 \%$ confluent 24 well plate and it corresponded to the Final Count (FC) of the first passage after transfection. Subsequent cell counts were used to plot growth curves for each clone, as displayed in figure 3.5.B. Simultaneously to the isolation of clones, the remaining cells derived from transfection were pooled together and considered as different populations (figure 3.5.A). As WI38 cells were thawed at $\approx$ PD 19 and 3 PDs were needed to expand the cells prior to transfection, in all graphs on figure 3.5 , the Y intersection on the X axis corresponds to PD 20 , indicating the total number of PD that WI38 had undergone before transfection. The best fit curves (sigmoidal) show that the log-phase (where cells were actively dividing) lasted until, approximately 30 cumulative days and was followed by the plateau phase (where cells divided more slowly as the percentage of senescence increased). The non-linear regression $R^{2}$ values for populations are 0.96 and 0.98 for shMSH2 and control populations respectively and KolmogorovSmirnov test indicated that the curves were not significantly different to each other ( $p=0.96$ ). The best fit curves for clones had an $R^{2}$ of 0.92 and 0.88
for shMSH2 and control clones respectively and Kolmogorov-Smirnov test also indicated that the curves were not significantly different to each other ( $p=0.60$ ), suggesting that depletion of the MSH2 protein in WI38 cells did not have a dramatic effect on cell growth. Furthermore, as clones behaved very similar to populations, it was assumed that clonal variation did not significantly impact on cell growth.

Figure 3.6.A shows the growth curves for the clones that were subsequently analysed for telomere length accounting for their MSH2 protein level (described on section 3.3.3 and shown on figure 3.9), and figure 3.6.B shows the decrease in the rate of cell division per day of those clones from approximately 0.8 PD/day after transfection to $0.2 \mathrm{PD} /$ day at the end of the experiment. Linear regression curves in figure 3.6.B (with $\mathrm{R}^{2} 0.92$ and 0.86 for control and MSH2 depleted clones respectively) exhibit significantly ( $p=0.0001$ ) different intercepts between both curves, indicating that MSH2 depleted clones divided more slowly than control clones; nevertheless, it was expected that senescence was reached at similar time points as slopes were not significantly different ( $p=0.6$ ) between controls and shMSH2 clones.


Figure 3.5: Clones and population growth curves. A summarises growth curves for populations and $\mathbf{B}$ for clones. The top graphs show individual growth curves and bottom graphs show the best fit curves (sigmoidal) for each counterpart on the top. For all graphs, red colours represent shMSH2-transfected cells and blue colours, controls.


Figure 3.6: Growth curves for clones analysed for telomere length. A plots the growth curves for the clones that were analysed for telomere shortening rates by STELA. The green points indicate the final PD at which cell pellets were taken for DNA extraction (the starting time points are not indicated because all fall within the same exponential region). B plots the rate of cell division as population doubling (PD) per day of the clones shown on graph A, grouped into controls (blue colour) and MSH2 depleted (red colour).

Cell viability was measured using Trypan blue staining during consecutive days and calculated by the equation 2.2. The percentage of viable cells was constant during the whole experiment (approx. $90 \%$ ) among all the samples (figure 3.7.A). The percentage of senescent cells was determined using the $\beta$-galactosidase assay as described in section 2.2.1.3. The senescence assay was conducted in duplicate and for some populations (e.g. shA. 3 or psupA.2), two different time points were assayed. As the percentage of senescent cells would increase with successive population doublings, the PD at which senescence was measured is indicated in the table below figure 3.7.B to allow comparison between clones. Figure 3.7.C summarises the percentage of viable (pink bars) and senescent (cyan bars) cells for the transfected WI38 clones and populations. Kruskal-Wallis tests showed no significant difference in viability and senescence (neither for the mean cumulative PD at which senescence was assayed) between control and shMSH2 clones or between control and shMSH2 populations. However, the percentage of senescent cells was significantly higher in shMSH2 clones than shMSH2 population, reflecting the effect of clonal variation.

Combining this data with the cell growth curves on figure 3.6 it was hypothesised that the lower cell division rate described in shMSH2 clones compared to controls might be a consequence of replication fork stoppage due to MSH2 depletion although no effect on viability was shown, suggesting that cells continued dividing and therefore, accumulating mutations that were not repaired by the MMR pathway. In addition, the fact that senescence was reached at similar time points for controls and shMSH2 clones suggests that less cell divisions occurred in the shMSH2 before entering senescence and therefore, an increase in telomere shortening rate might be anticipated in those clones.


Figure 3.7: Senescence and viability. A shows means $\pm$ standard errors (SE) for viability assay measured by Trypan blue staining. $\mathbf{B}$ shows means $\pm \mathrm{SE}$ for $\beta$ galactosidase assay. The number of cumulative PDs at which senescence was assayed and the number of cells scored is indicated in the table below the graph. The colour coding for A and B is: grey for non-transfected WI38 cells grown in parallel, red for shMSH2 and blue for controls; clear colour bars for clones and striped bars for populations. C summarises the percentages of viable cells (pink) and senescent cells (cyan) plotted on the left axis. Means and $\pm \mathrm{SE}$ are shown. Green dots represent the mean and SE of the cumulative PDs, plotted on the right axis, used for $\beta$ galactosidase assay. ${ }^{*} \mathrm{p}=0.015$ for one-way ANOVA uncorrected for multiple comparisons.

### 3.3.3 MSH2 downregulation in WI38 clones

To asses the level of MSH2 protein downregulation, western blots (WB) were optimised for protein concentration and primary antibody dilution using nontransfected WI38 cell pellets, as shown in figure 3.8, and reproducibility was assessed among independent protein lysates. Figure 3.8.A shows a linear relationship between the relative amount of MSH2 protein and the amount of primary antibody for 5 and $10 \mu \mathrm{~g}$ protein concentration that was lost for 20 $\mu g$ protein, as the increased protein concentration produced oversaturation of the control antibody. To avoid inaccuracies due to overexposure, $10 \mu \mathrm{~g}$ of protein and a dilution of $1 / 1,000$ for the primary anti-MSH2 antibody in $5 \%$ milk PBST (Phosphate Buffered Saline Tween-20) was used for subsequent experiments. To assess biological reproducibility, three different cell lysates (A, B and C) were assayed for the level of MSH2 protein expression. One-way ANOVA analysis with Tukey test correction for multiple comparisons showed that the protein content in lysates $A(p=0.03)$ and $C(p=0.002)$ was significantly different from B. Nevertheless, when using a centrifugation step of 10 minutes at $13,000 \mathrm{rpm}$ (in a 5415R Eppendorf Centrifuge) to remove cell debris (for cell lysates D, E and F), the variability of the relative MSH2 expression among biological replicates decreased to non significant levels (figure 3.8). Therefore, this centrifugation step was included before protein quantification by the Bradford assay (Bradford, 1976).

Figure 3.9 represents the relative expression of MSH2 to GAPDH (control protein), calculated using the normalisation method by the sum described in Degasperi et al. (2014), for clones and populations of WI38 cells transfected with either psuperior.puro control or psuperior.puro-shMSH2 plasmids. Briefly, the intensity of each band was divided by the sum of the total intensity for the same protein (MSH2 or GAPDH respectively) within a blot to obtain the relative amount of protein per blot, and subsequently, the relative amount of MSH2 for each lysate was divided by the total amount of GAPDH for the same lysate and blot, what allowed to compare replicates from different gels.

Clones sh143, sh123, sh110, sh101, sh112 and sh104 with a percentage of MSH2 downregulation of $99,94,93,63,54$ and 31 respectively were chosen


Figure 3.8: Preliminary western blots. A shows the relative amount of MSH2 measured by western blot using three different final protein concentrations ( 5,10 and $20 \mu \mathrm{~g}$ ) and three primary antibody dilutions ( $1 / 2,000,1 / 1,000$ and $1 / 500$ ) were assessed. B shows the relative amount of MSH2 was quantified in six lysates (A to F) from WI38 cell pellets collected at different time points and it is plotted as mean $\pm \mathrm{SE}$. For lysates $\mathrm{A}, \mathrm{B}$ and C no centrifugation step was performed after the lysis whereas for $\mathrm{D}, \mathrm{E}$ and F a centrifugation step of 10 minutes at $13,000 \mathrm{rpm}$ was performed. ${ }^{*} \mathrm{p}=0.03$ and $^{* *} \mathrm{p}=0.002$.
for subsequent telomere length analysis. Unfortunately, clones sh116, sh107 and sh109 with a greater than $50 \%$ protein downregulation were not suitable for telomere analysis as too small cell pellets at the end of the cell culture were obtained. In addition, clone sh100, having an $80 \%$ downregulation in a cell pellet corresponding to $\mathrm{PD}=48.6$, was not analysed for telomere shortening as another WB from PD $=43.5$ showed a greater MSH2 expression, suggesting that the MSH2 downregulation might have occurred as a later event. After transfection, the populations were heterogeneous for a) integration sites, b) copy number and c) plasmid expression, therefore, the level of the MSH2 protein was not significantly different from populations transfected with either control or shMSH2 plasmids and due to that heterogeneity, populations were not included in the telomere length analysis.


Figure 3.9: MSH2 downregulation in WI38 clones. The relative MSH2 protein expression in WI38 populations (A) or clones (B,C and $\mathbf{D})$ transfected with either control or shMSH2 plasmids is shown. Also WI38 non-transfected cells were assayed as a controls (shown in B). Each bar represents the mean of at least 4 independent measurements (up to 6-7 when sample was available) $\pm$ SE. Red refers to shMSH2 (labeled "sh") and blue to controls (labeled "cont."). For clones control 110, control 128 and sh 100, two cell pellets corresponding to two different time points were assessed, represented as $-s$ and $-m$ for start and middle time points respectively. $\mathbf{E}$ summarises the relative MSH2 protein levels for all the clones. Data from all control clones was pooled into one data set (blue) and the percentage of MSH2 downregulation was calculated comparing the mean of each clone to the mean of the pooled control clones. pvalues for one-way ANOVA uncorrected for multiple comparisons are shown for each clone. Clones selected for telomere analysis are indicated with a green arrow.

### 3.3.4 Telomere shortening in WI38 $\mathrm{MSH}^{+/-}$clones

Telomere length was measured using single telomere length analysis (STELA) (Baird et al., 2003) as described in section 2.2.2.8. When the telomere length is measured at two different time points with a known number of cell divisions (PDs) between them, the Telomere Shortening Rate (TSR) can be calculated as shown in equation 3.1, where mTL is the median telomere length for each time point and PDs is the number of population doublings separating the two time points.

$$
\begin{equation*}
\mathrm{TSR}=\frac{\mathrm{mTL}_{1}-\mathrm{mTL}_{2}}{\mathrm{PDs}} \tag{3.1}
\end{equation*}
$$

Four control clones (control100, control110, control112 and control128) and six MSH2 downregulated clones (sh143, sh123, sh110, sh101, sh112 and sh104) were analysed for three different telomeres (XpYp, 12q and 17p) using STELA. An example of four STELA gels for clones sh123, sh143 and control112 is shown in figure 3.10.

It has been demonstrated using FISH (Londoño-Vallejo et al., 2001) and STELA (Baird et al., 2003) that different telomere alleles might have distinct mean lengths and even distinct attrition rates ((Britt-Compton et al., 2006) and (Graakjaer et al., 2006)). Based on SNPs in the flanking (subterminal) sequences (Baird et al., 2000), the WI38 cell line is heterozygous for the XpYp (haplotypes A and B) and 12q (haplotypes B and $\Delta$ ) telomeres. Furthermore, it contains an allele in 17 p that can be amplified using the 17 p 6 primer designed in Britt-Compton et al. (2006). Therefore, haplotype specific primers were used for $12 q$ and XpYp to investigate allele-specific telomere shortening rates (for primer sequences refer to appendix B).

Figure 3.10: STELA blots and scatter plots for WI38 clones. The top part of the figure shows four blots for three telomeres ( $17 \mathrm{p}, 12 \mathrm{q}$ and XpYp). Two blots show both alleles (B and $\Delta$ ) for $12 q$ and the universal primer for XpYp shows both alleles (A and B). In all gels, a control DNA (extracted from the normal lymphoblastoid cell line CEPH 1375.02) was amplified with XpYpE2 primer to be able to compare and pool data from different gels. The quantification was done using ImageQuant software and HR ladder and 1 kb ladder were used for size calibration. The telomere length of each molecule (amplicon) was adjusted by substracting the flanking region where each primer anneals ( 3078 bp for $17 \mathrm{p} 6,231 \mathrm{bp}$ for $12 \mathrm{qnull} 3,544 \mathrm{bp}$ for $12 \mathrm{qSTELA}, 406 \mathrm{bp}$ for XpYpE2 and 434 bp for both allele specific XpYp primers). Scatter plots shown below each STELA blot were plotted using GraphPad Prism6. Median values and interquartile ranges are shown. For the DNA dilutions that produced no PCR bands (e.g. control 112 end), the DNA concentration was increased in subsequent STELA reactions.

Figure 3.11 shows four scatter plots for the telomere length of four clones (two controls and two shMSH2) where the two different alleles for the XpYp telomere were amplified. When the universal XpYpE 2 primer (that amplifies A and B alleles) was used, the mTL at the last time point was greater than the mTL at the first time point, impeding to accurately calculate TSR and indicating the presence of two different length alleles that were amplified at distinct efficiencies. Therefore, the use of haplospecific primers allowed us to study allele A and B independently and all four graphs show the presence of both telomere length subpopulations when the universal primer was used. Figure 3.12 shows the scatter plots for all clones analysed for each telomere in all 10 clones and table 3.1 summarises the median telomere lengths (mTLs) and telomere shortening rates (TSRs). Due to the limited amount of DNA, the haplotype-specific primers for XpYp were only assayed when the presence of two subpopulations with very different telomere lengths made the TSR estimate inaccurate.


Figure 3.11: Scatter plots for the XpYp telomere. In all graphs, blue indicates first and orange second, time points. When both XpYp alleles are shown, universal primer XpYpE2 was used for STELA. For A allele (Hap.A), the XpYp-427G/415C and for B allele (Hap.B), the XpYp-427A/415T primers were used.


Figure 3.12: Scatter plots for telomere length of WI38 transfected clones. A plots medians and interquartile ranges of three telomeres ( $\mathrm{XpYp}, 12 q$ and 17 p ) for four controls clones. $\mathbf{B}$ plots medians and interquartile ranges of three telomeres for six MSH2 downregulated clones. For all graphs each telomere is measured at the beginning of the cell culture (blue colour) and at the end (orange colour).

|  | control 100 | control 110 | control 112 | control 128 | sh 143 | sh 123 | sh 110 | sh 101 | sh 112 | sh 104 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| MSH2 downregulation | - | - | - | - | 99 \% | 94 \% | 93 \% | 63 \% | 54 \% | $31 \%$ |
| Population Doubling ${ }^{\dagger}$ | 5.12 | 9.66 | 6.00 | 13.77 | 2.94 | 3.55 | 5.23 | 4.52 | 8.77 | 6.98 |
| mTL-1 17p (\# values) | 2,461 (67) | 3,135 (31) | 2,073 (56) | 1,852 (22) | 2,176 (70) | 1,781 (27) | 2,528 (22) | 1,584 (29) | 2,580 (30) | 2,382 (51) |
| mTL-2 17p (\# values) | 2,388 (5) | 2,268 (42) | 2,005 (25) | 854.3 (26) | 1,742 (28) | 1,608 (44) | 2,365 (23) | 1,296 (36) | 2,059 (30) | 1,653 (33) |
| TSR 17p | 14.26 | 89.75 | 11.33 | 72.45 | 147.62 | 48.73 | 31.17 | 63.72 | 59.41 | 104.44 |
| mTL-1 12qB (\# values) | 5,976 (62) | 5,666 (40) | 4,876 (56) | 5,525 (98) | 5,261 (65) | 6,055 (29) | 5,997 (49) | 5,008 (81) | 5,557 (32) | 4,639 (68) |
| mTL-2 12qB (\# values) | 5,579 (17) | 4,817 (107) | 4,414 (17) | 3,456 (125) | 3,988 (24) | 4,665 (47) | 4,874 (20) | 4,456 (55) | 5,041 (18) | 4,270 (90) |
| TSR 12qB | 77.54 | 87.89 | 77.00 | 150.25 | 432.99 | 391.55 | 214.72 | 122.12 | 58.84 | 52.87 |
| mTL-1 12q $\Delta$ (\# values) | 5,630 (30) | 6,678 (30) | 5,462 (54) | 5,127 (24) | 4,891 (51) | 4,393 (12) | 5,082 (28) | 2,113 (13) | 5,282 (28) | 6,648 (28) |
| mTL-2 12q $\Delta$ ( $\#$ values) | 5,311 (7) | 5,744 (17) | 4,750 (5) | 3,954 (27) | 4,793 (30) | 4,212 (31) | 4,796 (18) | 1,704 (16) | 4,381 (46) | 6,132 (12) |
| TSR 12q $\Delta$ | 62.31 | 96.69 | 118.67 | 85.19 | 33.33 | 50.99 | 54.69 | 90.49 | 102.74 | 73.93 |
| mTL-1 XpYpE2 (\# values) | 9,643 (52) | 2,443 (85) | 835.9 (192) | 1,202 (171) | 8,904 (44) | 2,838 (37) | 10,382 (43) | 5,474 (97) | 6,105 (54) | 5,002 (216) |
| mTL-2 XpYpE2 (\# values) | 8,967 (24) | 2,071 (137) | 7,722 (18) | 2,308 (183) | 8,472 (27) | 2,902 (51) | 10,334 (31) | 4,676 (87) | 5,652 (47) | 7,623 (32) |
| TSR XpYpE2 (bp/PD) | 132.03 | 38.51 | * | * | 146.94 | * | 9.20 | 176.55 | 51.65 | * |
| mTL-1 XpYp Hap.A(\# values) | - | - | 698.3(144) | 6,094 (25) | - | 12,451 (11) | - | - | - | 5,202 (114) |
| mTL-2 XpYp Hap.A (\# values) | - | - | 570.9 (11) | 5,038 (47) | - | 11,517 (11) | - | - | - | 4,392 (12) |
| TSR XpYp Hap.A | - | - | 21.23 | 76.69 | - | 263.10 | - | - | - | 116.05 |
| mTL-1 XpYp Hap.B (\#values) | - | - | 8,938 (22) | 1,038 (38) | - | 2,568 (6) | - | - | - | 7,469 (27) |
| mTL-2 XpYp Hap.B (\# values) | - | - | 8,581 (11) | 739.7 (4) | - | 2,543 (8) | - | - | - | 4,910 (26) |
| TSR XpYp Hap. B | - | - | 59.50 | 21.66 | - | 7.04 | - | - | - | 366.6 |
| Combined TSR (bp/PD) | 71.5 | 78.1 | 57.6 | 81.9 | 190.2 | 152.3 | 77.5 | 113.2 | 68.2 | 86.8 |
| $\pm$ SEM | $\pm 24.3$ | $\pm 13.4$ | $\pm 19.5$ | $\pm 20.5$ | $\pm 85.3$ | $\pm 74.7$ | $\pm 46.7$ | $\pm 24.3$ | $\pm 11.7$ | $\pm 14.4$ |

Table 3.1: Summary of STELA results. Median telomere lengths (mTL) expressed in base pairs (bp) at two different time points (mTL-1 represents first time point and mTL-2, second), were obtained from the scatter plots on figure 3.12. ${ }^{\dagger}$ indicates the number of population doublings between the two time points. The telomere shortening rate (TSR) per telomere was calculated using the equation 3.1. The number of population doublings used per clone is also shown in the table and was calculated using the equation 2.1. An asterisk (*) next to the TSR indicates that no TSR was calculated based on the universal primer and allele-specific primers were used instead. "-" indicates that the allele-specific primer was not used for the specific clone. The combined TSR represents the mean for all telomeres measured in each clone $\pm$ SE. and will be used in figure 3.17.

Figure 3.14.A shows the mean for the $\mathrm{TSR} \pm$ SE per clone. ANOVA analysis showed no significant differences between control and shMSH2 clones due to the great error. Nevertheless, as the percentage of MSH2 downregulation decreased, the TSR variation between different telomeres was smaller as shown in figure 3.14.B, implying that downregulation of the MSH2 protein might have an impact on TSR, but it might be different for each telomere. To address this question, the TSR between controls and shMSH2 clones was plotted per telomere in figure 3.14.C and with the exception of the $12 \Delta$ allele, all telomeres in shMSH2 clones shortened at a higher, although not significant, rate compared to control clones, suggesting that some ends might be more vulnerable than others to the loss of the MSH2 protein. Therefore, the TSR was plotted per telomere accounting for the relative amount of MSH2 protein (figure 3.13) and it was found that the B allele for the $12 q$ telomere had an exponential correlation ( $\mathrm{r}=0.86$ ) with the level of MSH2 protein downregulation while the A allele of the XpYp telomere had a linear $(\mathrm{r}=0.96)$ correlation. Conversely, the other allele for the XpYp telomere had almost no relationship ( $\mathrm{r}=0.07$ ) with the level of MSH2 protein downregulation. Interestingly, the $12 \mathrm{q} \Delta$ allele showed a mild significant $(\mathrm{r}=-0.63)$ correlation between the TSR and the protein level although in the opposite direction than 12 qB allele.

To try to explain the different behaviour observed between the two alleles of the $12 q$ telomere, the telomere variant repeat region was studied into detail. As shown in figure 3.15.A, 17p and XpYp telomeres had a very homogeneous array of TTAGGG repeats. Conversely, both alleles for the 12 q telomere had many more interspersed repeats. In the telomere maps showed in figure 3.15.A, each letter refers to a hexanucleotide (e.g T=TTAGGG). Specific hexanucleotide repeat combinations that were present in both alleles or only in one of them were studied and private repeat combinations for $12 \mathrm{q} \Delta$ (CG, TTT, CTT, GTTG, TTGT, GGTT, TTTG and TGTT) and for 12qB (CC, TC, CT, CCC, TCC, CCCC and TCCC) were found. Those repeat combinations might be responsible for the different behaviour of both telomeres but in addition, shared repeats were more or less abundant in different alleles (e.g. GT, TT, TTG, GGT, GGTG and GTG were much more abundant in the
$12 q \Delta$ allele, compared to GG, GGG, TGT or TGTG that were more abundant in the 12 qB allele). It is also important to mention that there were 5 different di-hexanucleotide repeat combinations for the $12 q \Delta$ allele and 7 for the $12 q B$ allele. Taking into account that short repeats have a higher mutation rate than longer ones (reviewed in Ellegren (2004) and Shah et al. (2010)), the mutation likelihood of the 12 qB allele was expected to be higher confirming our results on telomere shortening rates, if a tendency towards deletions vs insertions occurred as a result of MMR deficiency (as suggested in chapter 4).


Figure 3.13: Relationship between TSR and the level of MSH2 downregulation. The TSR for control clones with no MSH2 downregulation is plotted in $\mathrm{X}=0$ and each shMSH2 clone, with their respective MSH2 protein level on X-axis. The shortening rate for each telomere is colour coded (red for 17 p , green for $12 \mathrm{q} \Delta$ allele, pink for both alleles of XpYp -assayed with the universal primer-, brown for A allele of XpYp and purple for B allele). Exponential lines are plotted per telomere with $R^{2}$ values of $0.74,0.40,0.08,0.002,0.93$ and 0.005 for $12 q B, 12 q \Delta, 17 p, \mathrm{XpYp}$, XpYp A and XpYp B respectively.


Figure 3.14: Telomere shortening rates in WI38 transfected clones. A plots the mean TSR and SE for all 10 clones analysed. A summary table for these values can be seen in table 3.1. B represents the TSR for all telomeres analysed (with a different colour code). The TSR grand mean for each clone is represented as an horizontal line. $\mathbf{C}$ is a bar graph with mean $\pm$ SE that compares the TSR for each telomere between controls (solid colour) and shMSH2 clones (stripped bars).

12qB


 (left), trinucleotides (middle) and hexanucleotides (right).

Telomere length, specially the length of the shortest and longest telomere, has been associated to replicative senescence and chromosome stability (Hemann et al., 2001) and it has been demonstrated that telomerase preferentially elongates the shortest telomeres in a primary cell line transfected with hTERT (Britt-Compton et al., 2009a). The initial mTL for the clones characterized here varied between 698 to $12,451 \mathrm{bp}$ among different telomeres and clones (see table 3.1 for more information). Due to that variability and taking into account the previous studies, I wanted to asses whether the initial telomere length had an effect on the TSR. Figure 3.16 is a plot of the starting mTL of each telomere against the TSR for control and shMSH2 clones. The graph also shows that telomeres with a longer starting mTL had a higher TSR compared to telomeres with a shorter starting mTL independently of the level of MSH2 protein. The linear regression analysis showed a slope of $0.01 \pm$ 0.003 and $0.02 \pm 0.009$ for controls and shMSH2 respectively but not significantly different to each other (pvalue $=0.67$ ), suggesting that downregulation of MSH2 did not alter the linear relationship between the starting mTL and the TSR. However, the $R^{2}$ values for the linear regression were 0.4 and 0.1 for controls and shMSH2 respectively, indicating that the heterogeneity in the TSR found in shMSH2 clones was greater than for controls, possibly due to the different levels of MSH2 downregulation.

### 3.3.5 MSH2 deficiency, TSR and oxidative stress

Free radicals produced during oxidative stress react with the purinic and pirimidinic bases within the DNA producing several types of DNA damage (Dizdaroglu, 2012). To protect against genomic instability, these lesions are usually repaired by any of the following DNA repair mechanisms: base excision repair (BER), nucleotide excision repair (NER) or MMR (reviewed in Bridge et al. (2014)). Within the genome, there are regions that are more vulnerable than others to oxidative stress depending on their nucleotide composition. von Zglinicki et al. (1995) and von Zglinicki (2002) showed that growing fibroblasts under hyperoxic ( $40 \%$ oxygen) conditions accelerates TSR from 90 bp to 500 bp per PD compared to fibroblasts grown in normobaric ( $21 \%$ ) oxy-


Figure 3.16: Effect of starting mTL on TSRs. The graph shows TSR in for four control (circles) and six shMSH2 (stars) clones for all different starting mTL. Each telomere is presented with a different colour: 17p in red, 12q B allele in blue, $12 q \Delta$ allele in green, XpYp in pink (both alleles), A allele in brown and B allele in purple. Regression lines are performed for controls (blue) and shMSH2 (red) clones with $R^{2}$ of 0.5 and 0.1 respectively. The dotted circles indicate that each telomere was clustered around a specific starting median length equally for controls and shMSH2 clones; the red circle highlights 17 p telomere clustered around $2,000 \mathrm{bp}$ and the other circle clusters both alleles for $12 q$ telomere around $5,000 \mathrm{bp}$.
gen conditions; this might be due to the fact that single strand DNA (ssDNA) breaks are induced at telomeres under oxidative stress ((Petersen et al., 1998) and (Honda et al., 2001)). The presence of extracellular superoxide dismutase in human fibroblasts slows the process of telomere shortening, extending the life span (Serra et al., 2003). In fact, BJ fibroblasts (normal human foreskin fibroblasts), which have a very low level of peroxide content, have a longer lifespan and a lower TSR of 15 to $20 \mathrm{bp} / \mathrm{PD}$ (Lorenz et al., 2001) compared to other fibroblasts with a higher peroxide content.

To confirm whether the reduction of oxidative stress impacts on the TSR when MSH2 is depleted, results from a previous study in which a primary cell line with depleted MSH2 was grown at $21 \% \mathrm{O}_{2}$ were compared to results obtained here. Figure 3.17 shows a comparison of the combined TSR for controls and shMSH2 clones at two different oxygen concentrations. Taking into account that oxidative stress increases the TSR (von Zglinicki et al., 1995), when comparing results from the current experiment (at $5 \% \mathrm{O}_{2}$ ) with that carried out by Mendez-Bermudez \& Royle (2011) (at $21 \% \mathrm{O}_{2}$ ), it would be expected to find a lower TSR for the control clones, due to the reduction of oxidative
stress in the present study. Surprisingly, one-way ANOVA analysis showed no significant differences between the TSR for controls under different oxygen conditions what might be due to a) primary cell specific telomere attrition rates or b) no changes on TSR between 21 and $5 \% \mathrm{O}_{2}$. Notwithstanding, the TSR obtained under $5 \% \mathrm{O}_{2}$ condition for WI38 cell line agreed with previous studies on the same cell line and oxygen conditions ((von Zglinicki et al., 1995) and (Dumont et al., 2001)). Two-tailed $t$-test showed significant differences ( $\mathrm{p}=0.005$ ) between controls ( $51 \pm 6 \mathrm{bp} / \mathrm{PD}$ ) and all shMSH2 clones ( $201 \pm 34$ bp/PD) for Mendez-Bermudez \& Royle (2011), contrasting with no significant differences ( $p=0.13$ ) between controls ( $72 \pm 5 \mathrm{bp} / \mathrm{PD}$ ) and shMSH2 clones $(115 \pm 20 \mathrm{bp} / \mathrm{PD})$ analysed in this chapter. However, the variances between both groups were significantly different $(p=0.03)$ as predicted by the greater heterogeneity in the TSR for shMSH2 clones. In addition, when clones with a greater than $60 \%$ downregulation level were compared with control clones, a moderate significance ( $p=0.05$ ) was shown. Finally, the significant difference found between shMSH2 clones grown at 21 and $5 \% \mathrm{O}_{2}$ might indicate that telomeres in MMR-deficient cells are more vulnerable to oxidative stress and hence, some of the high shortening found previously might be a consequence of accumulation of incorrectly repaired oxidative damage at telomeres.


Figure 3.17: Telomere shortening rates under different oxygen conditions. Comparison of the combined telomere shortening rates between this study and the one performed in Mendez-Bermudez \& Royle (2011). A shows mean $\pm$ SE per clone from Mendez-Bermudez \& Royle (2011) in grey (control) and black (shMSH2) bars and data from table 3.1 in blue (control) and red (shMSH2) bars. The level of MSH2 downregulation is shown below each bar for shMSH2 clones. B shows mean $\pm$ SE for all controls and shMSH2 clones from graph A. C shows same controls as B but for shMSH2 only clones with a greater than $60 \%$ protein downregulation are shown. One-way ANOVA p values are: ${ }^{* *} \mathrm{p}=0.01,{ }^{* * *} \mathrm{p}=0.001$ and ${ }^{* * * *} \mathrm{p}<0.0001$.

### 3.4 Discussion

In this experiment, stable clones with depleted MSH2 protein were generated and grew under low oxygen conditions ( $5 \% \mathrm{O}_{2}$ ) to study the effect of MSH2 downregulation on telomere shortening rate (TSR) with reduced levels of oxidative stress. In total, 21 clones for the WI38 primary cell line stably expressing the shRNA shown on figure 3.3 were isolated, 9 of which had a protein downregulation level higher than $50 \%$ demonstrated by western blot. In addition, 7 control clones expressing the plasmid without the shRNA were isolated in parallel. Kansikas et al. (2014) assayed the repair efficiency of human cells with variable proteins levels ( 25 to $75 \%$ ) for the most common MMR genes to mimic Lynch syndrome (LS) heterozygous patients and found that for MLH1 and MSH2 genes, the total loss of repair was reached when less than $25 \%$ of the protein was expressed. This suggests that only clones with a greater than $75 \%$ protein downregulation (in this study clones sh143, sh123 and sh110) might be expected to have completely depleted DNA mismatch repair function.

The telomere length was analysed by STELA (single telomere length analysis) and the TSR for three telomeres was calculated using the equation 3.1. The difference in TSR between control and all shMSH2 clones grown in 5\% $\mathrm{O}_{2}$ was not statistically different although a greater heterogeneity in TSR in MSH2 downregulated compared to control clones was observed. There was a trend towards an increase in telomere attrition rate for some telomeres (17p, $12 q B$ and XpYp A and B alleles) but not for the $12 q \Delta$ allele and specific repeats in the proximal telomeric region of this allele, that might be more vulnerable to deletions (CC, TC, CT, CCC, TCC, CCCC and TCCC) than others (CG, TTT, CTT, GTTG, TTGT, GGTT, TTTG and TGTT) were identified. Furthermore, although the linear relationship between the TSR and the starting telomere length was weaker for MSH2 than control clones, the linear trend was maintained, suggesting that shorter telomeres shortened at a slower rate than longer ones independently on the MSH2 protein level. Then, results shown in this chapter (under $5 \% \mathrm{O}_{2}$ ) were compared with those performed un-
der normoxic ( $21 \% \mathrm{O}_{2}$ ) oxygen conditions (Mendez-Bermudez \& Royle, 2011) and surprisingly it was found that the telomeres for control clones grown at different oxygen concentrations did not shorten at different rates. This might be explained by any of the following hypothesis:

1. Telomeres for the WI38 cell line might shorten at a higher rate than for the CCD34-Lu cell line at the same oxygen condition and therefore, under different conditions, telomeres might shorten at similar rates. This possibility is supported by the fact that the TSR for the WI38 cell line compared to other primary fibroblast strains (such as MRC5, IMR91 or HCA2) is greater (Britt-Compton et al., 2006), being $132 \mathrm{bp} / \mathrm{PD}$ the rate at $21 \% \mathrm{O}_{2}$ and $72 \mathrm{bp} / \mathrm{PD}$ at $5 \% \mathrm{O}_{2}$.
2. The effect on telomeres between cells grown at $21 \%$ or $5 \% \mathrm{O}_{2}$ is not significant to show differences between the TSR for controls. This possibility is unlikely as previous studies showed differences on shortening rates for other cell lines comparing those two oxygen concentrations (von Zglinicki, 2002), but as some primary cell lines are less vulnerable than others to oxidative stress (due to their high antioxidant content) such as BJ fibroblasts (Lorenz et al., 2001), it might be possible that the differences between the two cells lines in different oxygen environment is too low to be detected (or more telomere ends should have been tested to see an effect).
3. The TSR for the CCD34-Lu cell line might be underestimated if telomerase was elongating the shortest telomeres. In Mendez-Bermudez \& Royle (2011), transfection with hTERT was done to allow the primary cell line undergo a higher number of cell divisions before entering senescence. Although telomere shortening was observed despite the presence of telomerase, it might have occurred that the TSR reported was an underestimation because of the presence of telomerase. Being that the case, it would also be anticipated that the TSR for shMHS2 clones may also have been underestimated.

TSR comparisons between control and MSH2 downregulated clones presented in this chapter were only significantly different when clones with a greater than $60 \%$ downregulation were included. Furthermore, the TSR between MSH2 clones were significantly different between the two oxygen conditions, suggesting that the significant differences found at $21 \% \mathrm{O}_{2}$ arose, at least partially, as a consequence of an incorrect repair of oxidative damage at telomeres. ATR is known to suppress telomere fragility (reviewed in Verdun \& Karlseder (2007)) and is activated by replication stress (reviewed in Yang et al. (2004) and Zeman \& Cimprich (2014)) or different DNA damages, including oxidative stress ((Wang \& Qin, 2003) and reviewed in Yan et al. (2014)). In addition, MSH2 has been shown to recruit ATR after exposure to different damaging agents such as MNNG (Methylnitronitrosoguanidine) or cisplatin (Pabla et al., 2011) suggesting that the resistance to some chemotherapeutic drugs such as 5 ' FU (5-fluorouracil) in MMR-deficient colorectal patients might be due to the inability to activate the DNA damage Response (DDR) by MSH2-ATR interaction. Besides, reduced amounts of MSH2 might interfere with the correct repair of replication errors that are more prone in repetitive regions such as telomeres, increasing the telomere attrition rate if there is a tendency to accumulation of losses against gains of repeats. It has been demonstrated in mice that loss of MSH2 leads to telomere capping defects (Campbell et al., 2006) contributing therefore to telomere shortening and our results suggest that depletion of MSH2 contributes to an accelerate TSR partially due to a) an incorrect repair of oxidative damage at telomeres and b) an accumulation of uncorrected deletion loops that might be telomere-variant repeat specific.

## 4 Telomere instability in a cancer cell line deficient for MSH2

### 4.1 Background

The DNA mismatch repair pathway (MMR) is known to play a major role in correcting replication errors (single base or small insertion/deletion loops) that arise despite the selectivity and proofreading functions of DNA polymerases (reviewed in McCulloch \& Kunkel (2008) and Flood et al. (2015)). The first report of spontaneous mutation rates of a dinucleotide in normal human cells suggested that they occurred ranging from $3.1 \times 10^{-8}$ to $44.8 \times 10^{-8}$ mutations/nucleotide/generation (Boyer \& Farber, 1998) and more recently, a background rate of $5 \times 10^{-5}$ slippage mutations/locus/cell division was reported (Lever \& Sheer, 2010). As previously explained in section 1.1.4, mutation rates significantly increase in MMR deficient cancer cells (specially MSH2 ${ }^{-/-}$and MLH1 ${ }^{-/-}$) compared to MMR proficient (Hanford et al., 1998). This results in microsatellite instability (MSI), which has been proposed to be an early event in tumorigenesis in MMR defective cancers and persists after transformation (Shibata et al., 1994). Recently, Supek \& Lehner (2015) showed that mutations arising after MMR inactivation are spread homogeneously along euchromatic and heterochromatic regions in comparison with mutations in MMR-proficient tumours (that accumulate preferentially in heterochromatic regions), confirming that MMR defects increase genomic instability. Furthermore, Tomasetti et al. (2015) showed that the number of mutations in colorectal cancers (CRC) with MMR deficiency were about ten times higher than in MMR-proficient CRC, although he suggested that only three driver mutations were sufficient to trigger carcinogenesis using a mathematical model that considered the cancer incidence and the number of somatic mutations in MMR-deficient patients. Interestingly, higher mutation rates were also found in blood from LS patients (heterozygous carriers for a germline

MMR gene mutation) compared to blood from age-matched sporadic CRC patients ((Parsons et al., 1995) and (Coolbaugh-Murphy et al., 2010)), suggesting that the instability driven by the MMR defect might also occur, at lower levels, in normal tissues such as the germline, bone marrow or intestinal crypts (all tissues with a high mitotic index). If those mutant alleles were transmitted to the next generation through the germline, they might contribute to genetic anticipation as they might increase the cancer risk for an earlier onset (Seguí et al., 2013).

There are two types of mutations occurring in short tandem repeat (STRs) regions or microsatellites: point mutations and frameshift mutations, being the latter the most frequent (reviewed in Li et al. (2004) and Pumpernik et al. (2008)). The MMR pathway detects the single-base changes and the insertion/deletion loops correcting both types of mutations (Kolodner, 1996); however, under MMR-deficiency the frameshift mutations, occurring due to replication slippage, contribute to MSI. It has been shown that the genomic location, length of the array and sequence context affect the mutation rate of microsatellites (Harfe \& Jinks-Robertson, 2000b). The repeat unit size (mononucleotide, dinucleotide, trinucleotide, etc.) is inversely correlated to mutation frequency while motif length (number of units) is directly related (reviewed in Shah et al. (2010)).

Mutations in microsatellites have been proposed to affect tumour progression by three mechanisms: gene activation, gene inactivation or gene expression levels change (reviewed in Shah et al. (2010)). Actually, many protooncogenes such as $K R A S$ or $E G F R$ and some tumour supressor genes such as $R B$ or TP53, have repetitive regions within their sequence and in addition, cancer predisposing genes such as MSH2, MLH1, MSH6, PMS2, BRCA1, BRCA2, NF1, NF2, APC, PTCH1, and VHL also contain repetitive structures within their sequence (Centelles, 2012), (Schmidt \& Pearson, 2016) and (Smith et al., 2016). Besides, Ahuja et al. (1997) showed that MSI tumours had a higher promoter methylation frequency with a more extensive pattern due to an increase in the frequency of de novo CpG methylation, suggesting an interaction between the MMR pathway and chromatin remodeling activities.

Telomeres, due to their repetitive nature, might be expected to show high instability under MMR deficiency; in fact, Pickett et al. (2004) showed a higher telomere mutation frequency telomere mutation frequency (defined by the appearance of mutations that changed the sequence of the assayable proximal variant repeat region of the telomere) in MMR-defective colon tumours than in sporadic colon cancers and Mendez-Bermudez \& Royle (2011) suggested that this instability might be biased towards deletions, contributing to an increase in telomere shortening rate (as previously discussed in section 3). Furthermore, Shah et al. (2010) reviewed the mutational bias depending on the DNA structure and MMR defect and suggested that MSH2 and MLH1 mutants might have a bias to deletions. Among all MMR-defective colon cancer cell lines analysed in Bermudez (2007), LoVo ( $\mathrm{MSH}^{+/-}$) had the highest telomeric mutation frequency $16 \%$, compared to $M L H 1^{+/-}$(3.9-6.4\%) or $\mathrm{MSH}^{+/-}$( $5.6 \%$ ) cancer cells. A deeper look into the mutant molecules from the previous study also suggested that there might be a tendency towards deletions against insertions (5:1) although it was not statistically significant. Conversely, mutations in the hPMS2 gene have been involved in tetranucleotide repeat expansions ((TTTC/AAAG) and (TTCC/AAGG) sequences) suggesting that microsatellite instability due to different MMR gene mutations might have distinct impacts on telomeres (Shah \& Eckert, 2009) .

The traditional methods to measure MSI were based on PCR, allowing to amplify some mini or microsatellite markers in tumour and normal tissue and to distinguish between the mutated version and the non mutated based on their length ((Jeffreys et al., 1991) and reviewed in Geiersbach \& Samowitz (2011) and Kim \& Park (2014)). This has been used to discriminate between microsatellite stable (MSS) and unstable (MSI) tumours, contributing to the prognosis; however, due to the high frequency of mutated versus progenitor alleles in the tumour compared to the normal tissue, the frequency of the mutated allele for a specific loci cannot be estimated. The use of small pool PCR analysis can be used to detect the progenitor and the low-frequency mutant alleles thus, providing a more accurate estimate of the mutation frequencies (Coolbaugh-Murphy et al., 2004) and single molecule PCR has been used to
study germline mutations in microsatellites and expanded simple tandem repeats (Beal et al., 2015). However, MSI analysis based on PCR are usually restricted to the specific markers used for clinical diagnosis and give poor information about MSI across the genome. Nishant et al. (2009), Kim \& Park (2014) and Zavodna et al. (2014) reviewed how can the new sequencing technologies contribute to explore this issue as they allow to a) identify somatic alterations, b) estimate mutation rates, c) compare mutation rate variation between individuals and d) detect new rare microsatellite variants.

Baird et al. (1995) developed a PCR-based method to study the telomere variant repeats (TVR, described in section 2) and it has been used to estimate telomere mutation frequencies as mutational analysis on the canonical telomere repeat array cannot be performed ((Coleman et al., 1999), (Baird et al., 2000), (Pickett et al., 2004), (Mendez-Bermudez et al., 2009) and (MendezBermudez \& Royle, 2011)). However, it is important to note that a) pure repeat arrays are expected to have a higher slippage mutation frequency than short interspersed repeats and b) there is a limit in the type of mutations that can be scored by TVR, as only mutants that change the pattern (one single event) can be detected while mutations that restore the original pattern (more than one mutational event) will remain masked.

In this chapter, I present a XpYp telomere mutation analysis in the LoVo colon cancer cell line based on TVR analysis on single molecule STELA products to amplify individual telomeres. I compared the results with those shown in Bermudez (2007) to increase statistical power and to look for shared mutations. In addition, I exploited the potential of capillary electrophoresis to resolve TVR products to improve accuracy, resolution and throughput. Finally, I examined the use of a third generation sequencing technology (based on nanopores) to sequence the XpYp telomere in LoVo as a new approach to get a full map of the interspersion pattern.

### 4.2 Aims

Single molecule STELA (smSTELA) products have been previously used to detect somatic mutation frequency in normal and colon carcinomas (Jeyapalan et al., 2008), (Mendez-Bermudez et al., 2009) and (Mendez-Bermudez et al., 2012). Therefore, to understand whether MSH2 deficiency leads to gains or losses of repeats smSTELA products were generated from the LoVo ( $\mathrm{MSH}^{-/-}$) cell line to analyse telomere mutant molecules (defined by the appearance of mutations that changed the sequence of the assayable proximal variant repeat region of the telomere). The objectives for the aim described above were:

- To generate more than 100 smSTELA that will be further analysed by Telomere Variant Repeat (TVR) PCR.
- To resolve TVR products by capillary electrophoresis to increase throughput.
- To compare data with previous mutation analysis on the same cell line to look for shared mutants and to increase the sample size.
- To use a new third generation sequencing method (MinION ${ }^{\text {TM }}$ sequencing) to study STELA products from the LoVo cell line to explore the more distant telomere variant repeat region.


### 4.3 Results

### 4.3.1 Mutant telomeres analysed by traditional TVR

Generation of 303 smSTELA products was performed in 96 well plates by diluting LoVo DNA to $90-170$ pg per PCR. To confirm the presence or absence of PCR products and to measure the telomere length of the amplicon, size separation by agarose gel electrophoresis in a $0.8 \%$ LE agarose gel and subsequent Southern blot hybridisation to a random-primed $\alpha^{32} \mathrm{P}-\mathrm{dCTP}-l a b e l e d$ telomere and ladder ( 1 kb and HR) probes as described in section 2.2.2.8 was performed. Figure 4.1 shows an example of a blot with some single and double molecules and a diagram indicating primer annealing sites. Single telomeres generated by STELA were amplified in a final $20 \mu \mathrm{l}$ volume, and $5 \mu \mathrm{l}$ were run by agarose gel electrophoresis as shown in figure 4.1; therefore, less than 15 $\mu$ l were available for subsequent TVR-PCR (performed as described in section 2.2.2.9). Only $1 \mu$ was needed per PCR but different primer pairs and conditions were tested, and the abundance of negative results even for the positive controls, resulted in the final analysis of only 130 smSTELA products.


Figure 4.1: Single XpYp telomere molecules generated by smSTELA. The blot on the left shows some single and double bands that were run in a $0.8 \%$ agarose gel to screen for single bands. The LoVo cell line is heterozygous (haplotypes A and B) for the XpYp telomere and as only the A allele is informative for TVR, the allele-specific primer (XpYp 427G/415C) for this haplotype was used to generate smSTELA. The diagram on the right, shows the structure of the smSTELA products (green), indicating with a blue arrow the position of the primers used for STELA. The nested PCR that involves the analysis of the telomere variant repeats (TVR) is shown in red. The position of the forward primer (also haplotype A-specific) is shown with a blue arrow and all the potential binding sites for variant repeat-specific primers with grey arrows.

Application of the original protocol described by Baird et al. (1995) resulted in the detection of some (33) variant repeats maps within the smSTELA products but the abundance of negative results suggested that our conditions were
suboptimal. To control for PCR independent variables, e.g. gel mix and electrophoresis conditions, large amounts of 10X TBE and $6 \%$ acrylamide-urea gel mix were produced for exclusive TVR use. In addition, annealing temperatures tested by gradient PCR (shown in figure 4.2) and labeling conditions of the TS30T primer by dot blot (shown in figure 4.3) were assessed. A range of annealing temperatures (from 61 to $67^{\circ} \mathrm{C}$ ) was assayed with no significant change in the amplification efficiency from $63-66^{\circ} \mathrm{C}$ (figure 4.2). We decided to use the highest temperature $\left(66^{\circ} \mathrm{C}\right)$ as it seemed to be more stringent for the TagTelW primer that, despite being specific for TTAGGG repeats amplification, it can also amplify TGAGGG repeats under suboptimal PCR conditions. For the denaturing acrylamide gel electrophoresis conditions, 60 W were used and depending on the track to be resolved, longer ( $>3$ hours) or shorter ( 2 h $30^{\prime}$ ) time were employed.

In addition, the 5 ' end labeling conditions for the flanking primer (TS30T) were optimised as described in section 2.2.2.5. Briefly, a template for the subtelomeric XpYp telomere was dot blotted and the TS30T primer labeled using different conditions was used as probe. The method described in Baird et al. (1995) uses $0.5 \mu \mathrm{M}$ for the allele-specific forward primer (TS30T in our experiment) so we tested concentrations ranging from 0.2 to $1 \mu \mathrm{M}$ to discover if the primer concentration had an effect on labeling efficiency. The top part of figure 4.3 shows that concentrations above $0.6 \mu \mathrm{M}$ of TS30T primer gave similar and higher relative intensities than concentrations below $0.6 \mu \mathrm{M}$, suggesting that we could use the concentration recommended by Baird et al. (1995). Nevertheless, we further used different amounts of $\gamma^{32}$ ATP and found that reducing the amount of radiation (from 37 to $26 \mathrm{~Bq} /$ reaction) and increasing the primer concentration (from 0.5 to $1.0 \mu \mathrm{M}$ ) we could reach higher labeling efficiency (bottom part of figure 4.3).

A


Figure 4.2: TVR optimisation conditions. A is a denaturing acrylamide gel electrophoresed at 60 Watts (W) for 3 hours, 2 hours and 30 minutes and 2 hours to show the time effect on TVR product resolution. Two different smSTELA products (B7 and B8) were run and three annealing temperatures were used ( 61,62 and $63^{\circ} \mathrm{C}$ ). B shows two acrylamide sequencing gels comparing the effect of power and time. The one on the left, containing the same products that on figure A, was run at 55 W for 2 hours and 20 minutes and the one on the right, containing two new products (E2 and E7) was run at 60 W for 2 hours and 35 minutes. G represents TGAGGG repeats (amplified using TagTelX primer), T represents TTAGGG repeats (amplified using TagTelW primer) and J represents TTGGGG (amplified using TagTelJ primer) repeats.

Using the conditions described before (annealing temperature $66^{\circ} \mathrm{C}, 26$ $\mathrm{Bq} /$ reaction and $0.1 \mu \mathrm{M}$ of TS30T primer) and in the methods (section 2.2.2.9), 33 molecules were analysed by 5 ' $\gamma^{32} \mathrm{P}-$ dATP end labeled flanking primer, two of which were the same deletion mutant (mutant A ) involving the deletion of 6 T repeats and giving a mutation frequency of $7.7 \%$. Figure 4.4.A shows an example of a gel including one mutant molecule and 8 non mutants. Their


Figure 4.3: Primer labeling efficiency using dot blots. A shows quantification of dot blots (on the left) using different TS30T primer concentrations (ranging from $0.2 \mu \mathrm{M}$ to $1.0 \mu \mathrm{M}$, A to E respectively). B shows quantification of dot blots (on the left) using combinations of primer concentrations ( 1.0 and $0.6 \mu \mathrm{M}$ ) and $\gamma^{32}$ ATP concentrations ( 26 and 37 Becquerels/reaction). Mean and standard error of the mean from triplicate experiments is shown. ${ }^{*}$ p $=0.01$ and ${ }^{* * *} p=0.0003$ using Fisher's test.
corresponding telomere variant map is also indicated in figure 4.4.B and compared with that previously generated by Mendez-Bermudez \& Royle (2011). Although a very similar telomere map was obtained, the two more distant blocks of T repeats that were previously described could not be amplified; furthermore, many negative results (including positive controls) were generated and the repetition of experiments reduced the amount of smSTELA available for further analysis. In addition to those reasons, the method described in Baird et al. (1995) is low-throughput and labor-intensive and mutants are scored manually; therefore, the same principle was applied to capillary electrophoresis as described in section 2.2.2.9 to improve throughput.

A


NNNGGGGGGGGNTTTTMNTTTTTTNTTTTMNJJJJJJJJJJJJJJJJNNNNJJJJJJJNNNNNNNNNNNNJJJJJJ . . . NNNGGGGGGGGNTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJJNNNNJJJJJJJJNNNNNNNNNNNNJJJJJJ . . . NNNGGGGGGGGNTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJJNNNNJJJJJJJNNNNNNNNNNNNJJJJJJ . . . NNNGGGGGGGGNTTTTMNTTTTTTNTTTTNNJJJJJJJJJJJJJJJNNNNJJJJJJJJNNNNNNNNNNNNJJJJJJJ . . . NNNGGGGGGGGNTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJNNNNJJJJJJJJNNNNNNNNNNNNJJJJJJJ. . .
MutantA $\rightarrow$ NNNGGGGGGGGNTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJJNNNNJJJJJJJJNNJJJJJJJ. . .
NNNGGGGGGGGNTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJJNNNNJJJJJJJNNNNNNNNNNNNJJJJJJ . . . NNNGGGGGGGGNTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJJNNNNJJJJJJJNNNNNNNNNNNNJJJJJJ . . . NNNGGGGGGGGNTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJJNNNNJJJJJJJJNNNNNNNNNNNNJJJJJJJ . . . NNNGGGGGGGGNTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJJNNNNJJJJJJJJNNNNNNNNNNNNJJJJJJ . . .
B Map generated here
Map in Mendez-Bermudez, A
\& Royle, N., 2011


Figure 4.4: Telomere variant maps from LoVo DNA. A shows details for an example TVR gel and their related variant map. Each vertical lane represents amplification of a particular variant repeat and 10 different smSTELA are shown. Key for the map: G represents TGAGGG repeats (amplified using TagTelX primer), T represents TTAGGG repeats (amplified using TagTelW primer) and J represents TTGGGG (amplified using TagTelJ primer) repeats. B shows a comparison of the maps generated in this experiment with previous published data from the same MMR-deficient cell line. Black boxes indicate inconsistencies between the two maps.

### 4.3.2 TVR by capillary electrophoresis

We first validated the suitability of capillary electrophoresis for our purpose using PCR products on the DRr (right direct repeat) region of the Human Herpesvirus-6A (HHV-6A). This region was chosen because a) the HHV-6A genome contains telomere-like repeats allowing it to integrate into the human telomeres (Huang et al., 2014) and b) large amount of amplicons of a determined size could be generated. Amplicons were generated by Dr. Victoria Cotton with U100F and TJ1R primers on different cell lines and when this amplicon was assayed by TVR with TagTelW primer it produced different TVR patterns depending on the cell line of origin. Figure 4.5 compares the patterns obtained either using a $5^{\prime} \gamma^{32} \mathrm{P}-\mathrm{dATP}$ end labeled flanking primer or a 5' FAM-labeled TagTelW primer and resolving the PCR products in an ABI 3100 Genetic Analyzer (see section 2.2.2.9 for a detailed comparison of the two methods), indicating that capillary electrophoresis is not only capable of detecting TVR products and distinguishing between different patterns but it also gives a greater resolution as products longer than 1 kb were detected.

Subsequently, smTELA products for LoVo DNA were assayed using FAMTagTelW, NED-TagTelX and HEX-TagTelJ primers. To improve the signal intensity we increased the number of PCR cycles from 19 to 24 for all three primers and annealing temperatures were maintained as $66^{\circ} \mathrm{C}$. The use of fluorescence and capillary electrophoresis allowed us to detect the distant $T$ block that was not detected using 5 ' end labeled flanking primer and no other additional repeats were observed using capillary electrophoresis suggesting a) generation of the same map and b) absence of PCR artifacts due to the primer modifications. Chromatograms were analysed using the GeneMapper v4.1 software. A set of bins was created with a mean size for each bin corresponding to the location of the first nucleotide of each repeat $\pm 1 \mathrm{nt}$, allowing comparisons to the progenitor allele. Figure 4.6 shows a comparison of the progenitor allele map generated using 5 ' $\gamma^{32} \mathrm{P}$-dATP end labeled flanking primer and 5' fluorescent end labeled variant repeat-specific primers, indicating the position and correlation of each peak within the map.


Figure 4.5: Validation of TVR analysis by capillary electrophoresis. The top part of the figure shows five telomere maps generated by Dr. Victoria Cotton using the 5' $\gamma^{32} \mathrm{P}$-dATP end labeled flanking primer (TJ1R) and TagTelW reverse primer as described in Baird et al. (1995). The bottom part of the figure shows five chromatograms from the same amplicons using FAM-labeled TagTelW primer and the ABI 3100 Genetic Analyzer for resolution of the PCR products. Red lines show the comparison for each block. The chromatogram presents all amplification products following a linear relationship compared to the size ladder; however, the mobility of the products in a denaturing acrylamide gel follows an exponential, non-linear, function and therefore, to allow comparison of the results between both methods, the red lines shown on the figure are not straight. The codes BAN59, 7022, COR264, KK and 813 refer to each cell line.

A




B



C


Figure 4.6: Telomere variant map details of the progenitor $X p Y p$ A haplotype. $A$ is a screenshot of the GeneMapper v4.1 software used to screen TVR fragments. Vertical grey, blue and green lines are bins for G (TGAGGG), T (TTAGGG) and J (TTGGGG) repeats respectively. B shows a detail of the progenitor chromatogram generated by capillary electrophoresis. The top part of the figure shows three superimposed chromatograms (black for TGAGGG, blue for TTAGGG and green for TTGGGG repeats). Letters below each chromatogram indicate the letter and colour used to generate the telomere map. Colours for the map are different to those in the chromatogram to be able to compare the telomere map with previous versions in (Bermudez, 2007). C compares the map generated using either fluorescence or $\gamma^{32} \mathrm{P}$-dATP with the reference allele from the previous study. Black boxes indicate inconsistencies between the maps.

In total, 130 smSTELA were analysed by TVR (33 using $\gamma^{32} \mathrm{P}$-dATP and 97 using fluorescence and capillary electrophoresis), 25 of which contained mutations ( $19.2 \%$ mutation frequency) corresponding to 7 different types of mutations (see figures 4.7 and 4.9 for a detailed comparison of the mutant molecules with the progenitor allele):

- Mutant A: this mutant was present two times and showed a large deletion of the most distal block of J repeats. It was estimated that, at least, 18 repeats were lost, thus around 108 bp were lost per molecule due to this mutant.
- Mutant B: this mutant was present two times and contained the same mutation as clone 54 in Bermudez (2007). Two J repeats were lost in the first block of 15 J repeats, resulting in the detection of only 13 J repeats; this was equivalent to 12 bp lost per molecule.
- Mutant C: this mutant was present four times and contained two repeat deletions. As the 15 J block of repeats started one repeat earlier compared to the progenitor allele but no $T$ repeat lost was found, we estimated that one of the repeat loss corresponded to an N repeat between the third T block and the first J block. In addition, there was one T repeat loss in the 9 T block repeat, resulting in the displacement of the last $J$ block of repeats. In total, 12 bp were lost per molecule as in mutant B.
- Mutant D: this mutant was found only once and it showed lost of 2 J repeats, one in the first block of 15 J , resulting in a track of just 14 J , and the other in the second block of 7 J , resulting in a track of 6 J . As total number of base pair loss, this mutant was equivalent to B and C and 12 bp were loss.
- Mutant E: this mutant was found only once and it showed a gain of 1 J repeat in the 7 J repeats block, resulting in a block containing 8 J repeats and gain of 1 T repeat in the 9 T block. In total, 12 bp were gained in this molecule.
- Mutant F: this mutant was the most frequent and it was found 14 times. It contained a complex semi-duplication in the most distant portion of the 500 bp variant repeat array that could have arisen as a result of single strand annealing recombination (as shown by Tishkoff et al. (1997b) in $S$. cerevisiae when the $R A D 27$ gene was mutated). The complex insertion included: 6 G repeats, between the 7 J and the 6 J blocks, and following the last J block, $4 \mathrm{~T}, 2 \mathrm{~N}, 10 \mathrm{~T}$ and 15 J , resembling the variant map of the progenitor allele but in a more distant position. We estimated that, at least, 150 bp were gained in those mutants and due to its complexity and the abundance of molecules carrying this mutation, it was expected that it had occurred very early during the cell expansion. There was one insertion mutation in Bermudez (2007) that could not be described (clone 79) as the gain occurred further in the array. As this mutant was the most frequent, it might be anticipated that it might contained the same mutation as clone 79.
- Mutant G: this mutant was found only once and it contained a 4 repeat deletion. One of them, occurred in the first block of 5 T , where only 4 where detected; the next one, containing two J repeat deletion occurred in the 15 J block, what would be equivalent to mutant B , and the last one, occurred in the 9 T block where one T repeat resulted in the detection of only 8 T repeats. In total, 24 bp were lost in this molecule.

To calculate the mutation rate (probability of a mutation occurring per cell division) based on the Luria and Delbrück method (reviewed in Pope et al. (2008)), mutation data is needed at different time points to estimate the likelihood of a mutation event to occur earlier or later in the cell growth. However, the calculation of the mutation frequency (proportion of mutants present in a population), although less informative, only requires number of mutants. If the number of cells per tube is known, the fluctuation test would allow to calculate the mutation rate as the number of cells in the population approximately equals the number of cell divisions. In the present experiment, no data was obtained based on the total cell number and no estimates could be done on the number of cell divisions, thus, only mutation frequencies are reported.

The mutation frequency of $19.2 \%$ corresponded to the 25 mutant molecules, described above, from the total of 130 molecules analysed. However, if we considered it in a more conservative way, each different type of mutant (7 types) would represent a distinct mutation event and the abundance of one type compared to another, might only reflect how early or late it occurred during cell expansion and not how likely is this event to happen. Being that the case, the mutation frequency would be $5.4 \%$ as shown in table 4.1 and this hypothesis would explain why we found mutant F 14 times. As it is a very complex insertion in which, part of the variant map was semi-duplicated without interruption of the J repeats pattern, it might be expected to occur only once as an early event and not 14 independent times. Thus, considering that we found 5 different types of deletion mutants (all corresponding to an overall loss of 168 bp ) and only 2 different types of insertions (corresponding to an overall gain of 162 bp ), our results suggested that deletions were 2.5 times more frequent than insertions, confirming previous results in Bermudez (2007).


Figure 4.7: Details of the mutant XpYpA telomeres found in LoVo DNA. The progenitor allele is shown at the top of the figure. Four deletions found by capillary electrophoresis are shown below (the fifth mutant -A- was detected using $\gamma^{32} \mathrm{P}$-dATP). The two insertion mutants are shown at the bottom of the figure. The number of molecules scored per mutant are shown below the mutant name.

In addition, to prove that the same molecule was not analysed more than once (and therefore biasing the data), smSTELA products were sized using ImageQuant software, version 7.0. All smSTELA had different lengths indicating that distinct molecules were analysed and that the mutations found more than once (e.g. mutant F) reflected their frequency. Figure 4.8 shows the size for all progenitor allele in purple, with a median $\pm$ interquartile range telomere length of $2873 \pm 831 \mathrm{bp}$. In addition, the size of the molecules containing insertions or deletion is plotted in green and orange respectively. The median length $\pm$ interquartile range for the $F$ mutants was $2868 \pm 758 \mathrm{bp}$, not significantly different from the progenitor allele, suggesting that we were analysing the same population of molecules.


Figure 4.8: Size of the smSTELA analysed by TVR. The scatter plot shows the size of all 130 smSTELA analysed by TVR, differentiating between progenitor allele (in purple), deletion (in orange) and insertion (in green) molecules. For the progenitor allele and mutant F, medians and interquartile ranges are shown.

Subsequently, we compared telomere maps generated here (figure 4.9.A) with those from Bermudez (2007) (map comparison shown in figure 4.9) to a) increase the population size of mutant molecules and b) detect shared mutations. Only mutant B described here was shared between both studies (being the same mutation found in clone 54). Interestingly, as the same mutation was found in mutant G , accompanied with 2 T repeats deletions, we could hypothesise that the mutation $B$ arose earlier in time, and that, for mutant $G$, in successive $S$ phase cycles, it would have accumulated the additional loss of the 2 T repeats. Furthermore, in both studies it was found a deletion of one
$J$ repeat in the largest block of J repeats that was either accompanied by loss of 1 or 2 T repeats (clones 28 and 33 respectively) or a loss of an additional J repeat in the next $J$ repeats block (mutant $D$ ), suggesting that the loss of the first J repeat might have been another early and shared event in both studies. Additionally, one large deletion (mutant A) and one complex insertion (mutant F) were found in the present study but were not previously reported. We could hypothesise that mutant A originated relatively late during cell expansion, as only two molecules were found containing the same mutations; however, as discussed previously, mutant F was likely to have originated as an early event and therefore might have been expected to be detected in the previous study. Interestingly, one of the mutations found in Bermudez (2007) (clone 79) could not be characterised as the location of the mutation in the molecule impeded gel resolution, and as mutant F was very frequent and contained a duplication increasing the size of the array, we could hypothesise that clone 79 contained the same mutation.

The mutation frequency for the combined data based on the number of mutant molecules scored (40/223) was $17.9 \%$ with no significant differences between the two studies (Exact Fisher's test). Table 4.1 summarises the mutation frequency accounting for the number of molecules and the type of mutation observed in both studies and figure 4.10 shows a bar graph comparing the results. In both mutation analyses, a greater variation of deletion types was found in comparison to insertions (9:3); however, more insertion molecules were found in total in the present study compared to the earlier one. Therefore, the lack of significance between deletions (23/223) and insertions ( $16 / 223$ ) in the combined data (1.4:1 ratio) might be due to the fact that the same mutant molecules were detected more than once. Those mutants, specially mutant F, might be more frequent and were randomly analysed according to their frequency, thus, artificially increasing the mutation frequency. When only accounting for the type of mutant (13/223), the mutation frequency decreased to $5.4 \%$ in the combined data; furthermore, there were 9 different types of deletions versus 3 insertions resulting in a $3: 1$ ratio.
(bp/mut
N
(105)
$\begin{array}{ll}(2) & -108 \\ (2) & -12 \\ (4) & -12 \\ (1) & -12 \\ (1) & -24\end{array}$
Total (10) -168
7

+ JJJJJJJJ...(14) +150 Total (15) +162 (bp/mut \begin{tabular}{cc}
$\underline{\mathrm{N}}$ \& $\underline{(\mathrm{bp} / \mathrm{mut}}$ <br>
$(78)$ \& <br>
\& $(1)$ <br>
\& -12 <br>
\& $(2)$ <br>
\& -18 <br>
\& $(1)$ <br>
\& -12 <br>
\& $(6)$ <br>
\hline Total \& -6 <br>
\hline

 

$\underline{N}$ \& $\underline{(\mathrm{bp} / \mathrm{mut}}$ <br>
$(78)$ \& <br>
\& $(1)$ <br>
\& -12 <br>
\& $(2)$ <br>
\& -18 <br>
\& $(1)$ <br>
\& -12 <br>
\& $(3)$ <br>
Total \& -6 <br>
\hline

 

$\underline{N}$ \& $\underline{(\mathrm{bp} / \mathrm{mut}}$ <br>
$(78)$ \& <br>
\& $(1)$ <br>
\& -12 <br>
\& $(2)$ <br>
\& -18 <br>
\& $(1)$ <br>
\& -12 <br>
\& $(3)$ <br>
Total \& -6 <br>
\hline
\end{tabular}

$\begin{array}{ccc} & (1) & +30 \\ \text { Total } & (1) & +30\end{array}$
Total (1) +30
Figure 4.9: Telomere variant repeat maps comparison for the LoVo (MSH2 $2^{--}$) cell line. A shows the progenitor allele map at the top. Each letter, G, N, T and J, represents a 6 base pair

molecules are grouped by deletions or insertions. B shows data from (Mendez-Bermudez \& Royle, 2011) carried out in the same cell line as comparison. Maps for 6 of the 7 mutants found are
shown.
molecules are grouped by deletions or insertions. B shows data from (Mendez-Bermudez \& Royle, 2011) carried out in the same cell line as comparison. Maps for 6 of the 7 mutants found are
shown.
Arogenitor allele: NNNGGGGGGGGNTTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJNNNNJJJJJJJNTTTTTTTTTNNJJJJJJTTTTTTTTTTTTTTTT...
Deletions:
Mutant A: NNGGGGGGGGNTTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJNNNNJJJJJJJ....

 Mutant D: NNNGGGGGGGGTTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJNNNNJJJJJJNTTTTTTTTTNNJJJJJJTTTTTTTTTTTTTTTT...
 Insertions: Mutant Mutant F:NNNGGGGGGGNTTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJNNNNJJJJJJJNNNNNGGGGGGJJJJJJTTTTNNTTTTTTT Insertions:


> B Progenitor allele:


## Deletions:

Clone 28: NNGGGGGGGGGTTTTTNNTTTTTTTNTTTTNNJJJJJJJJJJJJJJNNNNNJJJJJJJTTTTTTTTTTTJJJJJTTTTTT... Clone 33: NNGGGGGGGGGTTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJNNNNJJJJJJJTTTMTTTTTJJJJJTTTTTT... Clone 54: NNGGGGGGGGGTTTTTNNTTTTTTNTTTTNNNJJJJJJJJJJJJJNNNNNJJJJJJJTTTTTTTTTTTTJJJJJTTTTTH... Clone 60: NNGGGGGGGGGTTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJNNNNJJJJJJJTTTTTTTTTTTJJJJJTTTTTT... Clone 64: NNGGGGGGGGGTTTTTNNTTTTTTNTTTTNNJJJJJJJJJJJJJJJNNNNNJJJJJJJTTTTTTTTTTJJJJJTTTTT...
Insertion: 5 repeat insertion

|  | Garrido, 2016 | Mendez, 2007 | Combined |
| ---: | :---: | :---: | :---: |
| Molecules analysed: n | 130 | 93 | 223 |
| Deletion molecules: $\mathrm{n}(\%)$ | $10(7.7)$ | $13(14.0)$ | $23(10.3)$ |
| Insertion molecules: $\mathrm{n}(\%)$ | $15(11.5)$ | $1(1.1)$ | $16(7.2)$ |
| Not defined: $\mathrm{n}(\%)$ | $0(0)$ | $1(1.1)$ | $1(0.5)$ |
| Total mutant molecules: $\mathbf{n}(\%)$ | $\mathbf{2 5}(\mathbf{1 9 . 2})$ | $\mathbf{1 5}(\mathbf{1 6 . 1 )}$ | $\mathbf{4 0 ( 1 7 . 9 )}$ |
| Types of del.mutants: $\mathrm{n}(\%)$ | $5(3.8)$ | $5(5.4)$ | $9^{*}(4.0)$ |
| Types of ins.mutants: $\mathrm{n}(\%)$ | $2(1.5)$ | $1(1.1)$ | $3(1.4)$ |
| Not defined: $\mathrm{n}(\%)$ | $0(0)$ | $1(1.1)$ | $1(0.5)$ |
| Total types of mutants: $\mathbf{n}(\%)$ | $\mathbf{7 ( 5 . 4 )}$ | $\mathbf{7 ( 7 . 5 )}$ | $\mathbf{1 2}^{\dagger}(\mathbf{5 . 4})$ |
| Deletions: bp involved | -168 | -60 | $-216^{* *}$ |
| Insertions: bp involved | +162 | +30 | +192 |
| Total balance of bp | -6 | -60 | -24 |

Table 4.1: Mutation frequencies for LoVo. n : number of molecules or mutant types; \%: percentage; del: deletions; ins: insertions. ${ }^{*}$ as mutant $B$ is same mutant as clone 54 , nine different types of mutants instead of ten were considered; ${ }^{* *}$ to avoid duplication, the 12 bp loss corresponding to mutant B and clone 54 were only considered once; ${ }^{\dagger}$ as discussed in the text, the not defined mutant might be mutant F so to avoid duplication, it has only being included once in the total type of mutants.


Figure 4.10: Graphical comparison of the mutation analyses in LoVo. A shows the total number of molecules analysed in brown and how many of them were deletions (purple) or insertions (orange). B shows in blue the total number of mutant molecules scored, splitting them in deletions (red) and insertions (green).

### 4.3.3 Sequencing telomeres using the MinION ${ }^{\text {TM }}$ device

The use of the TVR method allowed us to study the telomeric mutation frequencies based on the first 600 bp of the telomere variant repeat array but gave no further information about the telomeric structure following this repeat array. Due to their repetitive nature, telomeres are challenging structures to sequence and although Sanger sequencing was performed on reamplified smSTELA products using either XpYpE2 or Telorette2 primers, the presence of overlapped reads impeded to interpret results. I wanted to study the LoVo XpYp A allele, analysed in the previous section, in further detail to discover if more distant variant repeats were present in the telomeric sequence, so I could characterised this allele better for future analysis.

Telomeres are not suitable for short reads sequencing platforms as their repetitive structure impedes accurate assembly, but the development of third next generation sequencing technologies, that allow long reads sequencing, has become an exciting opportunity to explore the telomeric structure. There are only two commercially available platforms for long reads sequencing at the moment: single-molecule sequencing in real time (SMRT) developed by Pacific Biosciences ${ }^{\circledR}$ and nanopore sequencing using the MinION ${ }^{\text {тм }}$ device developed by Oxford Nanopore ${ }^{\text {TM }}$ Technologies. The SMRT technology has recently been used to sequence mice telomeres (Tong et al., 2015), demonstrating its suitability for sequencing long repetitive molecules, but, to our knowledge, the nanopore sequencing technology has not yet been explored for telomere sequencing.

The MinION ${ }^{\text {TM }}$ device only started to be commercially available in May 2015 but we had the opportunity to access it earlier through the MinION ${ }^{\text {тм }}$ Access Programme (MAP). The MinION ${ }^{\text {TM }}$ device is a portable sequencer ( 87 g of weight and dimensions: width 105, height 23 and depth 33 mm ) for long molecules, providing integral reads of single molecules loaded onto its flowcell. It contains a USB 3.0 port that will transfer the sequencing information from the consumable flow cell, connected to the device, to a host computer, where the MinKNOW software is run. Each flow cell contains up to 512 nanopores embedded in a polymer membrane and it integrates an ASIC
(Application-Specific Integrated Circuit) and a sensor chip. The former applies an electric potential across each nanopore and the latter measures the resulting ionic current flow. Each nanopore has four sensors (although only one is active at a time) detecting changes in the electric flow due to the interaction with the analyte (in our case, a deoxyribonucleotide but it can be adapted to detect ribonucleotides or amino acids). The sensor chip collects the information (at a rate of tens of thousand electric signals per second) from each of the active sensor and sends the information to the MinKNOW software that: a) receives the data, b) analyses it and sends feedback in real-time, c) controls the device selecting the run parameters, d) identifies and tracks the sample and e) ensures that the platform chemistry is performing correctly. Figure
4.11 shows a brief description of the device and its components.


Figure 4.11: MinION ${ }^{\text {TM }}$ device description. Figures obtained from Oxford Nanopore ${ }^{\mathrm{TM}}$ Technologies and composition generated for this Thesis for explanatory purposes. The MinION ${ }^{\text {TM }}$ device and the flowcell are shown on the left side of the figure. On the right side of the figure, a schematic representation shows a detail of the disruption in the current when a molecule pass through the nanopore.

The DNA library loaded onto the flowcell is double stranded, containing an anchorage molecule (a processive enzyme) in 5' called tether (shown in brown in figure 2.1). This enzyme a) binds to the nanopore, b) allows the DNA sequence to pass through the nanopore one base at a time in a single stranded manner and c) recognises a single stranded overhang (called leader) located at the 3 ' position of the DNA sequence, detecting when the complementary strand of the DNA molecule is being sequenced. Nucleotides are read as a
combination called K-mer generating a specific disruption in the current flow that can be used to determine the order of bases in a particular DNA molecule.

There are two workflows that can be used to base call the MinION ${ }^{\text {TM }}$ reads in Metrichor ${ }^{\mathrm{TM}}: 1 \mathrm{D}$ and 2D workflows. The former, generate reads from only one strand, and therefore are expected to be less accurate whereas the latter, generates three types of reads: "Template", "Complement" and "Two directions". The first two correspond to the forward and reverse strands while the latter consists of a consensus sequence created by the base calling software for each pair of template/complement (within the same molecule). Here, I will refer to 1D reads as those generated by the 1D workflow and I will include all three types of reads generated by the 2 D workflow in the 2 D reads. The output data from the MinION ${ }^{\text {TM }}$ device are .fast 5 files (an application of the HDF5 standard files) that needed to be submitted to the base calling software Metrichor ${ }^{\mathrm{TM}}$, a cloud-based data analysis service (available in: https://metrichor.com/s/) provided by Oxford Nanopore ${ }^{\text {TM }}$ Technologies prior to its analysis. The sequencing analysis were performed in SPECTRE (Special Computational Teaching and Research Environment) that provides a Linux environment enabling large data-set analysis. To convert .fast 5 to .fastq files, poretools was used. This toolkit is under development by Nick Loman and Aaron Quinlan and is publicly available in the public repository GitHub (https://github.com/arq5x/poretools). LAST (Local Alignment Search Tool) was used to align the reads to a consensus sequence. The options for the alignment were: -Q1 -r1 -q1 -a1 -b1, indicating that the input was . fast $(-Q 1)$, the match score was $1(-r 1)$, the cost of a mismatch was $1(-q 1)$ and the cost of a gap was set at $1+1 \mathrm{x}$ gap length ( $-\mathrm{al}-\mathrm{b} 1$ ) (for more information about this tool go to the manual page: http://last.cbrc.jp/doc/lastal.html). The aligned .maf file was converted to .sam using a python script developed by Martin C. Frith (maf-convert. py). The use of the package samtools allowed us to convert the .sam file into .bam and to get some information about the alignment (for more details about this tool go to the manual page: http://samtools.sourceforge.net) and subsequently, the alignment was visualized using the Integrative Genomic Viewer (IGV) that can be freely down-
loaded in https://www.broadinstitute.org/igv/ (from the Broad Institute). Finally, to verify the alignment, we retrieved the consensus sequence generated by IGV and performed a ClustalW alignment to the reference using the MacVector software, V11.1.1.


Figure 4.12: Pipeline for MinION ${ }^{\text {TM }}$ reads analysis. Blue boxed rectangles indicate bioinformatic packages and green ones scripts used. Names in italics are files that resulted from data analysis. Dotted lines indicate a filtering step that was only performed for the telomere libraries sequencing analysis as a first alignment was done on the telomere flanking region and a second was performed only on the reads that aligned to the reference.

As no previous experiments had been carried out in our laboratory, we tried to sequence two well-characterised amplicons that contained some homopolymeric interspersed repeats as controls together with each telomere sequencing experiment. Libraries for each amplicon were generated as described in methods (section 2.2.2.4) and the sequencing analysis was performed as summarised in figure 4.12.

Four sequencing libraries were prepared, two for the control sequences OL A31 and KUK A49 (donated by Enjie Zhang) and two for the XpYp telomere from the LoVo (MSH2 ${ }^{+/-}$) cancer cell line (generated by STELA as described in section 2.2.2.4). The OL A31 control amplicon is 4558 bp long, containing $40 \% \mathrm{G}+\mathrm{C}$ content while the KUK A49 control amplicon is 4439 bp long, containing $39 \% \mathrm{G}+\mathrm{C}$ content. Both amplicons contain homopolymeric regions and for both, a concentration of $\approx 500 \mathrm{ng}$ was used to prepare the sequencing library. The first telomere library (Tel1) was amplified using the universal XpYpE2 primer that amplified haplotypes A and B and the second telomere library (Tel2) was amplified using the allele-specific $\mathrm{XpYp} 427 \mathrm{G} / 415 \mathrm{C}$ primer that amplifies A allele (the same allele that I previously used for mutation analysis). The expected size for the STELA products for both telomere libraries was around 4 kb . The DNA concentration was higher for both control libraries compared to the telomere libraries as for the latter, the STELA products were generated using 250 pg genomic DNA and 25 cycles (as described in section 2.2.2.4) to avoid generating collapsed PCR products compared to the amplification carried out for the controls, that was performed using 10 ng of genomic DNA and 35 cycles as one size single amplicon was obtained.

I used the OL A31 library as internal control for Tel1 library, thus, I first loaded the Tel1 library onto the flowcell and sequenced it for 24 hours. As flowcells can be active up to 72 hours and different experiment can be run in the same flowcell if necessary, I added the OL A31 library on the day after loading Tel1. This allowed me to test the flowcell and to compare the performance of the two different types of libraries (one composed of one single size amplicon -control- and the other, containing different size STELA products). Subsequently, I used the KUK A49 library as an external control, therefore, it
was loaded onto a different flowcell that was only used to sequence this control library. This allowed me to test a) the sequencing accuracy without interference of other DNA molecules not related to the amplicon and b) to increase the sequencing time (up to 48 hours). The Tel2 library was then loaded onto an independent flowcell (like KUK A49) and performance of both sequencing experiments was compared.

Figures 4.13 and 4.14 show screen shots of the sequencing runs from the minKNOW (software used to sequence and process the reads from the MinION ${ }^{\text {TM }}$ device). When the first telomere library was loaded onto the flowcell, 495/512 pores were active, in comparison with 293/512 that were active after 24 hours when the control OL A31 sequence was loaded. The abundance of pores in salmon colour (figure 4.13.B) contrasts with the shortage of it in figure 4.14.B. Those are unavailable pores that can be activated in the future in the presence of DNA and will turn into green while sequencing; the fact that they were more abundant when the Tel1 library was loaded suggests that not enough DNA was available to activate them. However, when the OL A31 library was loaded, the percentage of unavailable (salmon) pores decreased, and the percentage of reading pores (green) increased compared to the Tel1 library. In addition, the histograms show that reads for the telomere library (Tel1) had a median size around 3 kb (according to agarose gel electrophoresis, the median length is approximately 4 kb ) while for OL A31 reads the median was greater ( $\approx 15 \mathrm{~kb}$ ), not agreeing with the expected product size ( 4558 bp ) and suggesting that some concatenated artifacts might being sequenced.

Table 4.2 shows the result of the stats script from poretools for the first sequencing experiment in which OL A31 and the first telomere library were run. This script was run on the raw .fast 5 files, retrieved from Metrichor ${ }^{\mathrm{TM}}$, before the alignment was performed. More 1D reads were obtained for the control sequence than the telomere library but conversely, more 2D reads were obtained for the latter. However, the N50 values were similar between 1D and 2D reads for each library ( 256 and 532 bp longer in 2D than 1D for OL A31 and telomere library respectively) and agreeing with the estimate product size for OL A31 (4558 bp). As different length products were present in

Tel1 library, the N50 values only suggested that the median of the sequenced products had an estimated length of $\approx 2000 \mathrm{bp}$. The minimum values (shortest read length) for the 1D reads from OLA A31 ( 23 bp ) and the 1D and 2D reads from Tel1 library ( 21 and 6 bp respectively) were very low compared to the 2D reads for the OL A31 (2273 bp), suggesting that either the sequence started to pass through the pore but quickly detached from it or that some artifacts produced a disruption of the current through the pore, creating a false positive. Either way, the quality values for those reads are expected to be very low compared to expected size reads, and thus filtration of the data for quality scores might be useful for eliminating them.

|  | OL A31 1D | OL A31 2D | Tel1 1D | Tel1 2D |
| :--- | :--- | :--- | :--- | :--- | :--- |
| total reads | 1234 | 204 | 262 | 4305 |
| total base pairs | 4376014 | 874549 | 187796 | 3733073 |
| mean (bp) | 3546.20 | 4287.00 | 716.78 | 867.15 |
| median (bp) | 3948 | 4353 | 247 | 231 |
| minimum (bp) | 23 | 2273 | 21 | 6 |
| maximum (bp) | 15101 | 8982 | 5458 | 139998 |
| N25 (bp) | 4400 | 4472 | 2814 | 4190 |
| N50 (bp) | 4106 | 4362 | 1911 | 2443 |
| N75 (bp) | 3724 | 4220 | 771 | 1136 |

Table 4.2: Poretools stats output on the first sequencing experiment in the MinION ${ }^{\text {TM }}$ device. N25, N50 and N75 refer to the length for which the 25,50 and $75 \%$ respectively of all bases in the sequences are in a sequence of length $\mathrm{L}<\mathrm{N}$ (definition from: https://www.broad.harvard.edu/crd/wiki/index.php/N50).

For the sequencing analysis, reads generated from the control OL A31 library were firstly aligned to the reference sequence ( 4558 bp ) resulting in a .bam file with 180 and 571 reads for 1D and 2D respectively. The consensus sequence of the alignment was retrieved from IGV and aligned again to the reference using ClustalW and a $95 \%$ identity was reached for the 2D reads (see alignment in appendix C and details of the alignment in table 4.4); however, the 1D reads alignment resulted in a very low (8\%) percentage of identity (alignment not shown). In addition, as the control library was run after the Tel1 library, we wanted to discard any interference from the telomere sequences with the alignment and in fact, no alignment was produced when either the telomere or flanking sequence (see below for further explanation) were used as consensus, suggesting that the reads produced when the control sequence was loaded corresponded exclusively to the OL A31 amplicon (possibly due to the higher concentration of this library in comparison with the


B


Figure 4.13: Screen shots for Tel1 library sequencing run. A shows the profile of accumulated reads after 24 hours. The top part shows data (such as ID, Asic and heat sink temperatures, or voltage) in reference to the flow cell and in the bottom part there is an histogram indicating the frequency of reads of indicated sizes. $\mathbf{B}$ shows the channels panel after a 24 hours run. The top part of the figure is shared with A, and the bottom shows each nanopore channel (identified by a number in white). Colours are: black=saturated (unable to make further reads), cyan= zero (no strand is being read at the moment but it is active for future reads), green=single pore or strand (pores that are currently reading) and salmon=unavailable (not reading or available at the moment but can be active in the future). The green triangle in the corner of some pores indicates the active sensor (out of the four that has each pore) that is active.


B


Figure 4.14: Screen shots for the control OL A31 library sequencing run. A shows the profile of accumulated reads after 24 hours. The top part shows data (such as ID, Asic and heat sink temperatures, or voltage) in reference to the flow cell and in the bottom part there is an histogram indicating the frequency of reads of indicated sizes. $\mathbf{B}$ shows the channels panel after a 24 hours run. The top part of the figure is shared with A, and the bottom shows each nanopore channel (identified by a number in white). Colours are: black=saturated (unable to make further reads), cyan = zero (no strand is being read at the moment but it is active for future reads), green=single pore or strand (pores that are currently reading) and salmon=unavailable (not reading or available at the moment but can be active in the future). The green triangle in the corner of some pores indicates the active sensor (out of the four that has each pore) that is active.
telomere one).
The telomere library was aligned with a consensus telomeric sequence of the XpYp (haplotype A) telomere containing up to 9 kb of telomeric repeats (TTAGGG) ${ }_{\mathrm{n}}$ that commenced with 407 bp of subtelomeric sequence (corresponding to the A allele) followed by a block of variant repeats generated based on TVR analysis. However, and due to the high number of repeats, the alignment generated a high number of false positives, resulting in 249 and 6777 reads (for 1D and 2D respectively) aligned to this telomeric consensus, producing very low identity ( 11 and $3 \%$ respectively) and suggesting that the alignment was artificially created by the amount of repeats and not by real identity. Therefore in a next step, the flanking, subtelomeric region ( 407 bp ) was used as a first consensus to filter data (as shown in figure 4.12). Figure 4.15 shows the pairwise alignment generated with ClustalW. Highlighted in yellow are the known SNPs that differed between A and B alleles. We used a consensus for the A allele sequence but our sequences corresponded to both alleles; therefore, if the highlighted SNPs in the consensus were not enough for allele discrimination during the alignment, a mismatch might be expected in those SNP positions when a B allele read aligned to the reference. The red circles (in figure 4.15) indicate the mismatched positions over the SNPs and the blue circles, mismatched positions in locations other than the SNPs. As expected, the identity was greater ( 98 vs $93 \%$ ) for 2D than 1D reads and the location of mismatches in the former was restricted to SNPs.

Finally, using the .bam files generated aligning against the flanking reference, a list of read identifiers (corresponding to the sequences aligned with the flanking sequence) was generated for the 1D and 2D reads and were used to filter the original .fastq file creating a new one that only contained the reads aligning to the flanking region. Subsequently, the consensus for the telomeric sequence described above was used again on the filtered .fastq files. Using this additional step, we: a) reduced the number of sequences that aligned to the consensus telomeric reference: from 249 to 123 for 1D reads and from 6777 to 592 for 2 D reads and b) improved the accuracy of the alignment: from 11 and $3 \%$ identity (for 1D and 2D reads respectively) to $88 \%$ (for
both 1 D and 2D reads). In addition, the pairwise analysis comparing 1D and 2D consensus was high ( $93.5 \%$ ) suggesting that both types of reads might be useful for sequencing. Unexpectedly, in this second alignment, the accuracy in the flanking region decreased as more mismatches (103/407 vs 26/407 and 105/407 vs 5/407 for 1D and 2D reads respectively) were generated. However, the alignment of 4800 bp reproducing large regions of TTAGGG repeats suggested that this technology might be explored for telomere sequencing. Unfortunately, none of the non-canonical variant repeats studied by TVR (TGAGGG and TTGGGG) were reproduced by this library, although the consensus generated indicated either $T$ or $G(K)$ for the positions where the TGAGGG variant was expected and either A or $G(R)$ for the positions where the TTGGGG variant was expected, suggesting that around $50 \%$ of the sequences might contain the correct nucleotide (visualised on IGV with an allele frequency threshold of 0.5 ). However, this low accuracy impeded precise analysis of the telomere sequence.

The XpYp B allele in the LoVo cell line has been reported to have a homogeneous TTAGGG telomere with lack of variant repeats (Bermudez, 2007), hence, to avoid misinterpretation of our results during the alignment, we decided to create a pure A allele library by using the XpYp427G/415C primer. In addition, we used a different control library (KUK A49) that was run in a different flowcell to check for sequencing accuracy between experiments, to analyse sequencing performance of the MinION ${ }^{\text {TM }}$ in homopolymeric repeats and to extend the sequencing time for each library (up to 48 hours) in order to get the maximum number of reads (specially for the telomere library).

|  | KUK 49 1D | KUK 49 2D | Tel2 1D | Tel2 2D |
| :--- | :--- | :--- | :--- | :--- | :--- |
| total reads | 15943 | 12612 | 2322 | 2589 |
| total base pairs | 39358603 | 23737063 | 3354462 | 3667959 |
| mean (bp) | 2468.71 | 1882.10 | 1444.64 | 1416.75 |
| median (bp) | 790 | 660 | 423 | 633 |
| minimum (bp) | 9 | 119 | 9 | 5 |
| maximum (bp) | 363496 | 8553 | 69710 | 64935 |
| N25 (bp) | 8880 | 4408 | 7693 | 4406 |
| N50 (bp) | 7467 | 4305 | 4270 | 3031 |
| N75 (bp) | 3503 | 3907 | 2023 | 1825 |

Table 4.3: Poretools stats output on the second sequencing experiment in the MinION ${ }^{\text {tM }}$ device. N25, N50 and N75 refer to the length for which the 25,50 and $75 \%$ respectively of all bases in the sequences are in a sequence of length $\mathrm{L}<\mathrm{N}$ (definition from: https://www.broad.harvard.edu/crd/wiki/index.php/N50).


Figure 4.15: ClustalW (v1.83) multiple sequence alignment for Tel1 library. $1 \mathrm{~d} . \mathrm{LoVo}$.E2 and $2 \mathrm{~d} . \mathrm{LoVo}$.E2 indicate 1D and 2D reads respectively from the first telomere library in which the E2 primer was used to amplify A and B alleles of the XpYp telomere in LoVo. Reference refers to the A allele flanking, subtelomeric region of the XpYp telomere. Yellow colour indicates SNP positions, red circles show mismatches within the SNP positions and blue circles mismatches outside the SNP positions. Gaps inserted=0; conserved identities=402 and 379 for 2D and 1 D respectively; Pairwise alignment mode: slow; pairwise alignment parameters: open gap penalty=10.0; extend gap penalty=0.1.

Table 4.3 shows the stats results for the external control and the second telomere library (Tel2), indicating that many fewer reads were generated for the STELA products than for the control sequence. However, the N50 and N75 indicated that for both libraries, the average length of most of the reads was within the estimated product size (specially for 2D reads). Figure 4.16.A shows screenshots for the flowcell used to sequence KUK A49 library and 4.16.B, for the flowcell used to sequence the second telomere library. The greater number of green pores in the first figure compared to the second, indicates that a higher number of pores were actively sequencing the KUK A49 but not the telomere library. As a consequence, fewer reads were obtained for the STELA products as shown in table 4.3. In addition, more black (dead) pores at the end of the KUK A49 sequencing suggest that this flowcell was more active. This might be influenced by the concentration and purity of the input DNA but also by batch to batch variations in the quality of the flowcells (Ip et al., 2015).

The KUK A49 library generated two .bam files containing 8540 and 128 sequences (for 1D and 2D reads respectively) aligning to the reference. When the consensus sequence generated in IGV was retrieved and aligned to the reference using ClustalW, a very high (99\%) identity for both types of reads was observed (full alignment shown in appendix D). The location of the mismatches was different in 1D and 2D alignments suggesting that the combined use of 1D and 2D reads when possible might increase accuracy.

Reads from the second telomere library were filtered using the flanking region (as previously described for Tel1 library) to remove alignment artifacts. The .bam files produced contained 62 and 44 sequences (for 1D and 2D respectively) and figure 4.17 shows the pairwise alignment of the consensus generated to the flanking region. The percentage of identity obtained was the same as for Tel1 library ( 93 an $99 \%$ for 1D and 2D respectively), although in this case, the three mismatches for the 2D alignment were not located in the SNPs but in other positions suggesting the presence of only one haplotype sequences. Besides, the mismatches in the 1D alignment were independently located in or out the SNP positions and were shared with those for the first te-


Figure 4.16: Screen shots of the channels panel for two flowcells. A is the flowcell used to run the control KUK A49 library. B is the flowcell used to run the second telomere library. In both figures, the pictures on the left show the flowcells 10 minutes after loading the library and the pictures on the right, the same flowcell after a 48 hours run (before stopping the run). Colours are: black=saturated (unable to make further reads), cyan= zero (no strand is being read at the moment but it is active for future reads), green=single pore or strand (pores that are currently reading), salmon=unavailable (not reading or available at the moment but can be active in the future) and yellow=multiple sequences around the pore (none of them is sequenced at the moment but the pore is active). The green triangle in the corner of some pores indicates the active sensor (out of the four that has each pore) that is active.
lomere library suggesting similar accuracy. After the filtration step, a second alignment was generated using the telomeric consensus sequence resulting in 493 and 924 1D and 2D reads respectively aligning to it. The same identity as for the telomere library 1 was reached ( $88 \%$ ), but a greater difference was observed between 1D and 2D reads ( 84.7 vs $93.5 \%$ ) suggesting that the mismatches generated in the alignment were different between the two experiments. Similarly to the alignment for Tel1 library, the accuracy in the flanking region decreased when the telomeric repeat consensus was used for the alignment although the 2D alignment for Tel2 library had less mismatches than for Tel1 library (87/407 vs 105/407). Despite, the non-canonical variants being poorly recognised, one TGAGGG (position 496 in the alignment) and one TTGGGG (position 856 in the alignment) repeats were sequenced in the 2 D alignment (see appendix E). To review all the telomere alignments, the four consensus sequences ( 1 D and 2 D for telomere libraries 1 and 2) were aligned together with the reference telomeric sequence using ClustalW to see whether the mismatches were shared and whether the accuracy increased or not. Appendix E shows the full alignment. A large region of TTAGGG repeats (from position 890 to $\approx 4350$ ) aligned to the consensus without quality lost with increasing length (in contrast with Sanger sequencing that loses resolution with longer fragments) suggesting that this technology might be suitable for telomere sequencing. Interestingly, in a more distant position (from 4550 bp until the end) the alignment decreased in accuracy and the presence of blocks of mismatches suggests that in this region non-canonical variants might be present. Besides, TACGGG (position 4547), CTAGGG (position 4564) or TCAGGG (position 4582 and 4786) repeats were sequenced despite not being present in the reference sequence, suggesting the presence of a more distant variant repeat region not previously described.

Finally, we used the Java based application Qualimap v2.2 developed by García-Alcalde et al. (2012) to extract statistical and graphical information from the .bam files (table 4.4 shows a summary). The duplicated reads rate had an average of 26.7 and $39.9 \%$ for the two control sequences (OL A31 and KUK A49 respectively). Unexpectedly, the average of duplicated reads for
the two telomere libraries was higher in the flanking compared to the telomere alignments ( $44.7 \mathrm{vs} 35.8 \%$ ) in spite of the higher repeat content of the latter. In addition, the high G+C content in all our alignments ( $>40 \%$ ) was not underrepresented in the sequencing as the $\mathrm{G}+\mathrm{C}$ content in the mapped reads was very similar to each reference, suggesting a comparable performance for $\mathrm{G}+\mathrm{C}$ and $\mathrm{T}+\mathrm{A}$ regions. With the exception of the alignments for the 1D reads from Tel1 library, a good mean coverage (over 60X) was reached for all the alignments, although a greater than the mean standard deviation in the alignments for the telomere sequence indicates that the coverage was lower in some regions. Interestingly, the high percentage of identity reported by ClustalW multiple alignment (with the exception of 1D reads from the OL A31 library) imply that less than $12 \%$ error was due to the mismatches (as ClustalW alignment did not insert any gap). However, when using the Qualimap software to look into detail at the . bam files generated using LAST alignment, we observed a much higher error rate (especially for all 1D compared to 2D reads and greater when using the telomere consensus compared to the flanking), suggesting that gaps were firstly introduced to increase the alignment accuracy by LAST and that they have a greater contribution to the error rate than the mismatches generated by ClustalW later on. This might be due to the score options chosen for the alignment as mismatches were scored equally as gaps; furthermore, the frequency of gaps in homopolymeric regions was below $29 \%$ in all alignments suggesting that they occurred more frequently in non-homopolymeric regions.

1. 2d.hapA vs. Reference

Aligned Length $=407$ Gaps $=0$
Identities $=404$ (99\%)

| 2d.hapA | 1 | TTGTCTCAGGGTCCTAGTGTGTCTGGAATTGGTGGGTTCTTGGTCTCACTGACTTCAAGA |  |
| :---: | :---: | :---: | :---: |
| Reference | 1 | TTGTCTCAGGGTCCTAGTGTGTCTGGAATTGGTGGGTTCTTGGTCTCACTGACTTCAAGA <br>  | 60 |
| 2d.hapA | 61 | ATGAAGACGCGGAACCTCGCGGTGAGTGTTACAGTTCTTAAAGGTGGCATGTCCGGAGTT | 120 |
| Reference | 61 | ATGAAGACGCGGAACCTCGCGGTGAGTGTTACAGTTCTTAAAGGTGGCATGTCCGGAGTT <br>  | 120 |
| 2d.hapA | 121 | TGTTTCTTCTGATGTTCAGATGTGTTCTGAGTTTCTTCTTTCTGGTGGGGTTGTGGTCTC | 180 |
| Reference | 121 | TGTTTCTTCTGATGTTCAGATGTGTTCTGAGTTTCTTCTTTCTGGTGGGGTTGTGGTCTC <br> $\star * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * ~(~) ~$ | 180 |
| 2d.hapA | 181 | ACTGGCTCAGGAGTGAAGCTGCAGACCTTTGCGGTGAGTGTCACAGCTCAGAAAGGCART | 240 |
| Reference |  | ACTGGCTCAGGAGTGAAGCTGCAGACCTTTGCGGTGAGTGTCACAGCTCAGAAAGGCAGT <br> ***************************************************************) | 240 |
| 2d.hapA |  | GTRGACवSAAAGAGTGAGCAGTAGCAAGATTTATTGCAAAGAGTGAAAGAACGAAGCTTC | 300 |
| Reference | $241$ | GTGGAC $*$ AAAGAGTGAGCAGTAGCAAGATTTATTGCAAAGAGTGAAAGAACGAAGCTTC | 300 |
| 2d.hapA | 301 | CACAGTATGGAAAGGGACCCCATTGGGTTGCCACTGCTGGCTCAGGCAGTCTGCTTTTAT | 36 |
| Reference | 301 | CACAGTATGGAAAGGGACCCCATTGGGTTGCCACTGCTGGCTCAGGCAGTCTGCTTTTAT <br>  | 360 |
| 2d.hapA | 361 | TCTCTAATCTGCTCCCTCCCACATCCTGCTGATTGGTCCACTTTCAG 407 |  |
| Reference | 361 | TCTCTAATCTGCTCCCTCCCACATCCTGCTGATTGGTCCACTTTCAG 407 *********************************************** |  |

2. 1d.hapA vs. Reference

Aligned Length $=407 \quad$ Gaps $=0$ Identities $=381$ (93\%)


Figure 4.17: ClustalW (v1.83) multiple sequence alignment for Tel2 library. 1d. hapA and 2d. hapA indicate 1 D and 2D reads respectively from the second telomere library in which the $\mathrm{XpYp} 427 \mathrm{G} / 415 \mathrm{C}$ was used to amplify the A haplotype of the XpYp telomere in LoVo. Reference is the same sequence in figure 4.15. Yellow colour indicates SNP positions, red circles show mismatches within the SNP positions and blue circles mismatches outside the SNP positions. Gaps inserted=0; conserved identities=404 and 381for 2D and 1D respectively; Pairwise alignment mode: slow; pairwise alignment parameters: open gap penalty=10.0; extend gap penalty=0.1.

|  |  | OL A31 |  | KUK A49 |  | Tel1 |  |  |  | Tel2 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 1D | 2D | 1D | 2D | 1D | 2D | 1D filtered | 2D filtered | 1D | 2D | 1D filtered | 2D filtered |
| Reference stats | Name | OL-A31 | OL-A31 | KUK-A49 | KUK-A49 | Flanking | Flanking | Telomere | Telomere | Flanking | Flanking | Telomere | Telomere |
|  | Length (bp) | 4558 | 4558 | 4493 | 4493 | 407 | 407 | 9003 | 9003 | 407 | 407 | 9003 | 9003 |
|  | GC \% | 40 | 40 | 39 | 39 | 49 | 49 | 50 | 50 | 49 | 49 | 50 | 50 |
|  | Mapped bp | 693383 | 2307166 | 35112966 | 483815 | 1577 | 31261 | 363593 | 1073682 | 24232 | 17547 | 920033 | 1821836 |
| . bam <br> stats | N.of reads | 180 | 571 | 8540 | 128 | 4 | 80 | 123 | 592 | 62 | 44 | 493 | 924 |
|  | Mapped reads | 100\% | 100\% | 100\% | 100\% | 100\% | 100\% | 100\% | 100\% | 100\% | 100\% | 100\% | 100\% |
|  | Paired reads | 0\% | 0\% | 0\% | 0\% | 0\% | 0\% | 0\% | 0\% | 0\% | 0\% | 0\% | 0\% |
|  | Read min | 2519 | 1957 | 1071 | 1731 | 362 | 323 | 641 | 371 | 327 | 348 | 375 | 417 |
|  | Read max | 4708 | 4779 | 4893 | 4412 | 423 | 427 | 4800 | 5030 | 438 | 427 | 5078 | 5395 |
|  | Read $\overline{\mathrm{x}}$ length | 3866 | 3953 | 4148 | 3724 | 401 | 386 | 3036 | 1865 | 392 | 383 | 1930 | 2038 |
|  | Dup. reads | 53 | 319 | 7410 | 43 | 0 | 58 | 51 | 373 | 40 | 31 | 258 | 640 |
|  | Dup. \%* | 29.4 | 55.9 | 86.8 | 33.6 | 0 | 72.5 | 41.5 | 63.0 | 64.5 | 70.5 | 52.3 | 69.3 |
|  | Dup. rate (\% ${ }^{\dagger}$ ) | 22.8 | 30.6 | 59.9 | 20.0 | 0 | 50.0 | 31.9 | 37.9 | 45.5 | 38.5 | 34.0 | 39.4 |
| Nucleotide Content | N.of A's | 169280 | 546151 | 8465570 | 110976 | 344 | 6689 | 71916 | 210883 | 5233 | 3754 | 181683 | 349916 |
|  | \% of A's | 25.3 | 25.3 | 25.0 | 24.4 | 22.9 | 22.7 | 20.3 | 20.6 | 22.8 | 23.0 | 20.7 | 20.0 |
|  | N.of C's | 157421 | 451599 | 7961103 | 90240 | 328 | 6409 | 67430 | 133174 | 5026 | 3463 | 116756 | 261583 |
|  | \% of C's | 23.6 | 21.0 | 23.5 | 19.8 | 21.9 | 21.8 | 19.0 | 13.0 | 21.9 | 21.3 | 13.3 | 14.9 |
|  | N.of T's | 179037 | 643645 | 9289873 | 143983 | 446 | 8733 | 103063 | 314000 | 6753 | 4809 | 26946 | 523354 |
|  | \% of T's | 26.8 | 29.8 | 27.4 | 32.0 | 29.7 | 29.7 | 29.0 | 30.7 | 29.4 | 29.5 | 30.7 | 29.9 |
|  | N.of G's | 162264 | 515626 | 8216887 | 110529 | 383 | 7603 | 112757 | 364148 | 5935 | 4265 | 308637 | 61691 |
|  | $\%$ of G's | 24.3 | 23.9 | 24.2 | 24.3 | 25.5 | 25.8 | 31.8 | 35.6 | 25.9 | 26.2 | 35.2 | 35.2 |
|  | N.of N's | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|  | GC \% | 47.9 | 44.8 | 47.7 | 44.1 | 47.4 | 47.6 | 50.7 | 48.7 | 47.8 | 47.4 | 48.5 | 50.2 |
| Coverage | Mean | 152.1 | 506.2 | 7815.0 | 107.7 | 3.9 | 76.8 | 40.4 | 119.3 | 59.5 | 43.1 | 102.2 | 202.4 |
|  | SD | 30.0 | 77.5 | 942.8 | 20.9 | 0.6 | 10.8 | 46.6 | 169.4 | 8.1 | 5.1 | 140.9 | 291.3 |
|  | MQ ( $\overline{\mathrm{x}}$ ) | 245.8 | 238.4 | 246.5 | 240.3 | 243.1 | 240.2 | 135.3 | 130.2 | 241.6 | 236.8 | 132.6 | 141.5 |
| Mismatch and indels | Error rate (\%) | 63.6 | 48.6 | 64.5 | 48.9 | 23.4 | 20.7 | 60.6 | 42.9 | 21.9 | 20.6 | 43.5 | 44.7 |
|  | Identity (\%) | 8.0 | 95.0 | 99.0 | 99.0 | 93.0 | 99.0 | 88.0 | 88.0 | 93.0 | 99.0 | 88.0 | 88.7 |
|  | Insertions | 17493 | 61873 | 913774 | 12832 | 62 | 977 | 11596 | 52229 | 897 | 421 | 47389 | 83979 |
|  | Deletions | 15061 | 83303 | 715664 | 16527 | 52 | 1296 | 5629 | 34428 | 932 | 851 | 28887 | 48806 |
|  | $\mathrm{N}_{\mathrm{n}}$ IDls (\%) | 24.5 | 29.0 | 25.3 | 29.2 | 22.8 | 24.7 | 20.5 | 21.8 | 22.5 | 28.5 | 20.4 | 23.4 |



 MinION ${ }^{\mathrm{TM}}$ are not Phred based so error calculation by the Qualimap might be inaccurate.

### 4.4 Discussion

In this chapter, mutation analysis on the A allele of the XpYp telomere in the LoVo (MSH2 ${ }^{+/-}$) colon cancer cell line has been performed and it was compared with results in Bermudez (2007) to understand the telomere instability driven by MSH2 deficiency. Furthermore, it was demonstrated that flourescently labeled primers can be used to perform TVR analysis and that the use of capillary electrophoresis improved the resolution as well as introduced the possibility of accurately measure each PCR product using a size standard ladder. In addition, the use of bins for the location of each variant repeat block accelerated the analysis process allowing a higher throughput.

Combining data from radioactive and fluorescent TVR, a similar mutation frequency to the previous study ( 19.2 and $16.1 \%$ respectively) was obtained although more insertion molecules ( $15 / 130$ vs $1 / 93$ ) were scored, reducing the percentage of deletions from 14 to $7.7 \%$. This was due to somatic mosaicism (reviewed in Campbell et al. (2015)) as the same mutation molecules were found more than once (especially an insertion mutation called mutant F, that was found 14 times). Nevertheless, the diversity and trend of the mutations was similar, as in both studies 5 different types of deletions were found but only three different types of insertions were found across both studies. Furthermore, one of the deletions found in both studies was shared (mutant B and clone 54) and the most frequent insertion mutant found here (mutant F) might represent the mutation that could not previously be mapped. Interestingly, a final balance of 9:3 (loss:gain) for the combined data suggests that deletions were three times more common than insertions; additionally, a total of 216 bp were lost due to all 9 deletion mutants while only 192 bp were gained due to the 3 insertion mutants, suggesting that despite less nucleotides lost per deletion (mean of 24 vs 64 bp ), the greater variety of deletion mutants compared to insertions would have contributed to the increase in the telomere shortening rate of human primary cells with depleted MSH2 protein as shown by Mendez-Bermudez \& Royle (2011) and in chapter 3 of this Thesis.

All MSI tumours have shorter telomeres than MSS tumours and in colo-
rectal cancer (CRC) cell lines, MSI has been strongly correlated with telomere shortening but not with telomerase activity (Takagi et al., 2000). Taking into account that telomeres in MMR-defective CRC cell lines shorten despite the presence of telomerase activity, it might be expected that inability of repairing slippage errors during replication would produce deletion mutants contributing to telomere shortening. However, as the cell divisions progress, accumulation of deletions, due to defective MMR, might lead to a significant decrease in the median telomere length and thus, very short telomeres might be more vulnerable to insertions. Then, there might be a balance between gains and losses of repeats in cancer cells while in primary cells, that cannot escape senescence, accumulation of deletions will increase the telomere shortening rate although the influence of senescence will impede that insertions occur as a late event. Some mathematical models reviewed in Ellegren (2004) called stepwise mutation models would agree with this hypothesis as they demonstrated that short microsatellites are more prone to suffer insertions whereas long microsatellite tend to accumulate deletions. Additionally, Conomos et al. (2012) found, in cells using the ALT mechanism for telomere length maintenance, that telomere variant repeats (especially TCAGGG) recruited nuclear receptors such as TR4 with greater affinity than TRF2 at telomeres, suggesting that the TVR region of the telomeres is essential for shelterin complex binding. Therefore, it might be anticipated that deletions in this region, reducing the number of variant repeats, would contribute to an accelerated telomere attrition rate due to unprotected, uncapped, telomeres. To demonstrate that MSI in CRC cells drives to losses of repeats when telomeres are long enough to keep shortening but to gains when telomeres reach a size limit, it might be interesting to analyse changes in the mutation profiles between two different time points as well as studying telomere length and shortening rates.

Subsequently, the same XpYp telomere allele analysed by TVR was sequenced using a third next generation sequencing device called MinION ${ }^{\text {TM }}$ from Oxford Nanopore ${ }^{\text {TM }}$ Technologies, that allows to sequence single long molecules, in an attempt to obtain full information from the whole telomere
array. This new technology has already been used to sequence a phage genome (Mikheyev \& Tin, 2014) and several bacterial genomes ((Quick et al., 2015), (Karlsson et al., 2015) and (Laver et al., 2015)). Although the error rate has been described to be very high compared to other sequencing technologies ( $\approx$ $38 \%$ ) those rates were calculated using the R6 chemistry and as the technology is evolving, lower error rates are being expected. Karlsson et al. (2015) reported a consensus accuracy, after mapping to the reference genome the most common nucleotide at each position, of $99.8 \%$ with a coverage of 60 X .

We used LAST as alignment tool and a consensus sequence generated based on the TVR profile (for two telomere libraries) and two reference sequences generated by Enjie Zhang using the Ion Torrent technology ${ }^{\text {TM }}$ (for two control libraries). We reached a high percentage of identity (95 and 99\%) for the two control sequences validating the sequencing method for accuracy. In addition, and after pre-aligning reads to a flanking region of 407 bp , we reached also high percentages of identity in the two telomere libraries. The first one, containing A and B alleles for the XpYp telomere, had a lower (12\%) but identical error rate, based on mismatches, than the 1D reads for the second library (containing only A allele). Furthermore, the 2D reads for this library reached an even lower error rate ( $\approx 10 \%$ ) based on mismatches, suggesting that the sequencing accuracy was very high. However, the error rate reported by Qualimap v2.2 was greater than the calculated based on ClustalW (that only accounts for mismatches), suggesting that insertion/deletions (IDLs) had an important impact on the sequencing error rate. This might be due to the score options chosen for the alignment but also to the different scale (not Phred) used as quality scores given by the MinION ${ }^{\text {TM }}$. Moreover, as the error based on IDLs decreased significantly when the two telomere libraries were aligned to the flanking region compared to the telomere, for future analysis it would be desirable to use the flanking region as an anchorage and not only as a filter for alignments against a tandem repeat reference. Alternatively, a more stringent score might be used for the gap penalty, trying to impede certain degree of misalignment. Interestingly, the percentage of IDLs in the two control sequences, containing homopolymeric repeat regions, was far less
frequent ( $\approx 27 \%$ ) than in non-homopolymeric sequences, suggesting that generation of gaps by the alignment software was not only restricted to tandem repeats.

The main aim of using the MinION ${ }^{\text {TM }}$ to sequence the XpYp telomere in LoVo was to map the whole array of telomere variant repeats (TVR) and to be able to sequence through the canonical TTAGGG tandem repeat array in its entirety. Unfortunately, the variant repeats analysed by TVR (TGAGGG and TTGGGG) were poorly sequenced suggesting that using this technology to sequence variant repeats of unknown telomeric ends might be challenging. However, the lack of mismatches in a wide region of $\approx 4 \mathrm{~kb}$ of TTAGGG repeats in comparison with a block of mismatches near the end of the sequence suggested that our consensus sequence might be incomplete and thus, a de novo assembler might improve the sequencing results. The high error rate of de novo assemblers (particularly in repetitive regions) makes these tools highly inappropriate; however, some authors have recently used "hybrid" approaches to overcome this issue ((Madoui et al., 2015) and (Goodwin et al., 2015)). In those experiments, long reads from nanopore sequencing have also been sequenced using short reads technologies, what has been used to correct errors in the long reads, increasing the accuracy of de novo assemblers. Unfortunately, we could not use these approaches as telomeres are not suitable for short read sequencing technologies, but with the application of those experiments, improvements in the accuracy might be made in de novo assemblers for future applications over long repetitive sequences. Finally, the G+C content is known to affect sequencing performance in other platforms and it is still unclear whether the MinION ${ }^{\text {TM }}$ is affected by high G+C content, although Laver et al. (2015) showed that high G+C sequences were underrepresented using R6 chemistry. We used a newer chemistry (R7.3 with the genomic DNA kit SQK-MAP004) and found that the G+C content of the mapped reads was very similar to the reference, suggesting that in our libraries the sequencing performance did not depend on the nucleotide composition. This, is a developing technology and many bioinformatic tools need to be developed in order to analyse complex regions such as telomeres or other repetitive sequences.

## 5 Effect of telomerase inhibition in an MSH2 ${ }^{-/-}$cancer cell line

### 5.1 Background

Human telomerase is a ribonucleoprotein that elongates telomeres in germline and immortal cells (Blackburn et al., 1989) and it is active in $80-90 \%$ of the tumours ((Kim et al., 1994) and (Shay \& Bacchetti, 1997)) allowing telomere maintenance despite the high cell turnover. In addition, tissues with a high proliferation index (male germ cells, activated lymphocytes and some stem cell populations) also have some levels of telomerase activity (reviewed in Cong et al. (2002)). In the case of the normal mucosa in the human colon, three stem cell compartments have been described to be involved in the regeneration of this tissue (reviewed in Clatworthy \& Subramanian (2001) and Sipos et al. (2012)) and telomerase activity has been found in normal as well as tumourous colon tissue (Kim et al., 1994). The main function of telomerase is to elongate telomeres to allow progressive cell divisions without loss of the proliferative capacity (Allsopp et al., 1992), although other non-canonical activities related with transcriptional regulation and metabolic reprogramming also impact on cancer cell survival (Low \& Tergaonkar, 2013). In general, telomeres in colorectal tumours shorten (Hastie et al., 1990), despite telomerase being active in most of the cases (Li et al., 1996) and (Shay \& Bacchetti, 1997), and a relationship between telomere shortening and microsatellite instability (MSI) was found in a group of 55 colorectal carcinomas (Takagi et al., 2000). However, no relationship was observed between telomerase activity and MSI (Vidaurreta et al., 2007), suggesting that in mismatch repair deficient cancer cells the predominant telomerase function might be other than telomere maintenance.

The study of telomerase inhibitors as potential therapeutical agents has resulted in several drugs targeting either the reverse transcriptase (hTERT)
or the RNA component (hTERC) of telomerase in an attempt to increase senescence and/or apoptosis in the tumour cells (reviewed in Rankin et al. (2008) and Phatak \& Burger (2007)). Damm et al. (2001) described a group of small chemical compounds that were able to inhibit telomerase activity in vitro, and to reduce tumorigenicity using xenograft mouse models. BIBR1532 is a small non-nucleosidic molecule that is able to inhibit endogenous or recombinant telomerase by affecting the processivity of, mainly, the long reaction products (Pascolo et al., 2002) and not by directly competing with the binding efficiency to the DNA. It was shown by Damm et al. (2001) that treatment of cancer cells with the telomerase inhibitor BIBR1532 lead to a cell proliferation arrest that was reverted if the treatment was removed. In addition, they suggested that this proliferation arrest might be due to the induction of a senescent phenotype that was shown to have an increased telomere erosion rate measured by TRF and dysfuctional telomeres measured by Q-FISH. In addition, Nakashima et al. (2013) proposed that cell death via the apoptosis mechanism might also be occurring as a consequence of telomerase inhibition in cancer cell lines.

Distinct concentrations of BIBR1532 have been found to be needed to inhibit telomerase activity; $\mathrm{IC}_{50}$ values for Damm et al. (2001) were $0.1 \mu \mathrm{M}$ whereas for Barma et al. (2003) were $5 \mu \mathrm{M}$ ( 50 fold higher). This difference was attributed to the purity of the telomerase enzyme to be inhibited; while the former used purified telomerase, the latter used crude cell extracts that might contain inhibitory substances reducing the effect of the BIBR1532 drug. $\mathrm{IC}_{50}$ values for El-Daly et al. (2005) on JVM13 cells (immortalized B-cell line from a patient with prolymphocytic leukemia) were $52 \mu \mathrm{M}$ and $\mathrm{IC}_{50}$ values for Pascolo et al. (2002) were in the order of $0.1 \mu \mathrm{M}$ for HeLa cell extracts and in the order of $100 \mu \mathrm{M}$ for recombinant telomerase indicating the wide range at which this drug has been tested. El-Daly et al. (2005) demonstrated that concentrations higher than $30 \mu \mathrm{M}$ of BIBR1532 had acute cytotoxic effects, within 72 hours after the treatment, on leukemia cells but not in normal hematopoietic stem cells. In addition, the antiproliferative effect of BIBR1532 was not dependent on telomerase activity or telomere length itself but on the
increase of discrete dysfunctional telomeres and the associated loss of TRF2.
The suitability of telomerase inhibitors for treating MMR-defective tumours is not confirmed yet. On the one hand, it has been hypothesised that the efficiency of the treatment might be related to the initial telomere length (as cells with longer telomeres will undergo more cell divisions before senescence or apoptosis triggered by short telomeres occurs), and thus, tumours with shorter telomeres ( $\mathrm{MMR}^{-/-}$) would be benefited from a telomerase inhibition treatment as they will reach senescence earlier. On the other hand, it has been shown that after telomerase inhibition of a MSH6 ${ }^{-/-}$cancer cell, telomere elongation occurred in a telomerase-independent, ALT-like mechanism (Bechter et al., 2004), therefore, further studies on telomerase inhibition in MMR-deficient cancer cells are needed. Additionally, synthetic lethality studies have been performed in MMR-deficient cancer cell lines to explore its potential as therapeutic approaches. Martin et al. (2010) showed that the inhibition of POLB and POLG genes by siRNA in MSH2 ${ }^{-/-}$and MLH1 ${ }^{-/}$cancer cells respectively, produced an increase in oxidative lesions, thus the presence of an additive effect for telomerase inhibition in MMR deficient cancer cells has been explored here.

### 5.2 Aims

To study the suitability of telomerase inhibition as a potential treatment in MMR-defective tumours by understanding the effect of telomerase inhibition on a human colon cancer cell line defective for the DNA mismatch repair (MMR) pathway compared to a MMR-proficient colon cancer cell line. The main objectives for this aim were:

- To treat two colon cancer cell lines, MMR-deficient (LoVo) and MMRproficient (SW480) with the telomerase inhibitor BIBR1532.
- To assess telomerase activity after the treatment with the telomerase inhibitor BIBR1532 compared to untreated controls.
- To study cell cycle dynamics of the cancer cells treated with the telomerase inhibitor using flow-cytometry analysis (FAC).
- To measure telomere length by single telomere length analysis (STELA) to study telomere attrition rates with and without the telomerase inhibitor treatment.


### 5.3 Results

### 5.3.1 Partial inhibition of telomerase using BIBR1532

The ability of BIBR1532 to inhibit telomerase was first tested in crude cell extracts from the two cancer cell lines described in table 5.1. Parsch et al. (2008) demonstrated a reduction in telomerase activity of $\approx 70 \%$ incubating cell lysates from the chondrosarcoma cell line SW1353 with $10 \mu \mathrm{M}$ of BIBR15 for 15 minutes on ice. Here, the cell extracts were incubated on ice with $30 \mu \mathrm{M}$ of BIBR1532 to try to get a higher reduction of telomerase activity measured by the telomerase repeat amplification protocol (TRAP assay).

|  | SW-480 | LoVo |
| :--- | :--- | :--- |
| Tissue | Colon | Colon |
| Sex | Male | Male |
| Donor age (years) | 50 | 56 |
| Tumour type | Duke's type B | Duke's type C |
|  | adenocarcinoma | adenocarcinoma |
| MMR | Proficient $^{\text {Tumour suppressor gene }}$ | $\mathrm{APC}^{-/-}$ |

Table 5.1: Cell lines information. Data acquired from the American Type Culture Collection (ATCC) website. For more information refer to chapter 2.

Telomerase activity was measured by the telomerase repeat amplification protocol (TRAP assay) as described in section 2.2.2.10 and semi-quantification was performed by measuring the relative intensity of the first PCR band compared to the internal control. Figure 5.1 shows that in the two adenocarcinoma cell lines a telomerase inhibition of 92.6 and $58.5 \%$ (for SW480 and LoVo respectively) was observed. This reduction was significantly different compared to untreated lysates ( $p<0.0001$ and $p=0.0186$ for Exact Fisher's test respectively) although being more modest for LoVo than SW480.

To examine whether treatment of cells in culture with similar concentration of BIBR1532 had an effect on short term cell growth dynamics, the two cancer cell lines were grown during approximately 10 days in Dulbecco's modified Eagle medium containing either 10 or $30 \mu \mathrm{M}$ of BIBR1532, measured cell number and viability using Trypan blue and compared them with the same cells grown in Dulbecco's modified Eagle medium and $0.3 \%$ DMSO (the


Figure 5.1: Telomerase downregulation in cell extracts. A is an example for a TRAP assay gel. The cell lysate loaded onto each lane is indicated at the top. "hi"= heat inactivated. B is a bar graph indicating the relative telomerase activity as a measurement of the relative intensity of the first PCR band ( $1^{\text {st }}$ PCR) to the internal control (IC) on figure A (measured using ImageQuant v. 7.0). Means and SE are shown for two replicate experiments. ${ }^{* * * *} \mathrm{p}<0.0001$ and ${ }^{*} \mathrm{p}=0.0186$.
equivalent and non-cytotoxic concentration to drug dilution) as control. No significant differences (Exact Fisher's test) were found for either of the cell lines between the controls and telomerase inhibitor treated cells in terms of cell growth dynamics suggesting that none of the concentrations had cytotoxic or cytostatic effects. Unexpectedly, as shown in figure 5.2, SW480 had very low percentage ( $0-2 \%$ ) of cell death for all three conditions (control and two BIBR1532 concentrations) but conversely, the LoVo cell line had a very high percentage of cell death including the non-treated control cells (40-60\%), what was also confirmed by FAC analysis as described in section 5.3.2. As no significant differences between the treatments and control for viability or cell growth were observed, it could be concluded either that the use of BIBR1532, up to $30 \mu \mathrm{M}$, in cell culture seemed to have no effect on short term growth for the two adenocarcinoma cells, independently of the effect on telomerase activity or that the drug was not effectively delivered to the cells.


Figure 5.2: Cell growth curves and cell death for short-term treatment with 10 and $30 \boldsymbol{\mu M}$ of BIBR1532. The graphs represent growth curves (triangles and straight lines) for SW480 (left) and LoVo (right) treated with two different BIBR1532 concentrations, 10 (dark green) and 30 (red) $\mu \mathrm{M}$ or with $0.3 \%$ DMSO as control (light green). Growth curves are plotted in the left axis and percentage of cell death measured by Trypan blue (circles and dashed lines) plotted in the right axis. When two measurements were available, mean $\pm$ standard error are plotted.

The extraordinary proportion of dead cells shown in figure 5.2 for LoVo has not previously been reported in this cell line and as it was observed for all conditions, it could not be concluded that it resulted as a consequence of the treatment. Furthermore, the cell was checked by western blot for MSH2 protein levels to validate its identity (see figure 5.3), the telomere variant
map was reproduced as shown in chapter 4 and previous experiments showed the presence of high levels of microsatellite instability (Bermudez, 2007), confirming its authenticity. Although there was no explanation for the very high percentage of cell death, it was a very significant feature distinguishing both cancer cell lines and it arose an important consideration about cell division and the calculation of cumulative population doublings (PD); as the mean of cell death was $\approx 50 \%$ for LoVo, half of the population should have divided twice for each passage (and not just once) to counteract the cell death before reaching confluency, and therefore, it might be expected that the number of PD would have been underestimated (see figure 5.5.B for an schematic representation).



Figure 5.3: Western blot for LoVo. The left part of the figure shows an example western blot result for the expression of MSH2 protein in LoVo compared with three different cell lysates from WI38. The right part of the figure is a bar graph quantifying the relative amount of MSH2 compared to GAPDH as described previously. Means and standard deviations are shown. WI38 1, 2 and 3 are three different cell pellets that were used for the WB optimization (see explanation in section 3.3.3).

### 5.3.2 Cell cycle dynamics of colon cancer cells treated with BIBR1532

Figure 5.2 shows that neither 10 nor $30 \mu \mathrm{M}$ of BIBR1532 had a detrimental impact on short-term cell growth and as I aimed to get the maximum telomerase inhibition but without cytotoxic effects (El-Daly et al., 2005), the two cell lines were cultured with $30 \mu \mathrm{M}$ for approximately 60 days to study telomere length and cell cycle dynamics. Growth curves are shown on figure 5.4 and linear regression analysis showed very high $\mathrm{R}^{2}$ values ( $>0.9$ ), indicating that, despite the treatment with the telomerase inhibitor, senescence was not reached as no stationary phase was detected. For treated SW480 cells, a slower growth rate was observed after passage 4 that was even more pronounced after passage 8 , resulting in significantly ( $p=0.0001$ ) different slopes between control and treated cells. Contrarily, in the LoVo cell line, treated and control cells grew at similar rate during 60 days, suggesting that BIBR1532 might have a cytostatic effect for long term treatment in SW480 but not in LoVo.


Figure 5.4: Cell growth dynamics of cancer cell lines treated with $\mathbf{3 0} \boldsymbol{\mu M}$ of BIBR1532. The graphs represent growth curves for SW480 (left) and LoVo (right) for control (light green) and $30 \mu \mathrm{M}$ BIBR1532 (red) treated cells.

To study cell cycle dynamics, cells were collected for each passage and assayed for DNA content using propidium iodide staining and flow-cytometry analysis (FAC), as described in section 2.2.1.6. All samples (from different passages, cells and treatments) were analysed by FAC at the same time to reduce batch and setting variations and duplicate samples were collected for


Figure 5.5: Cell cycle description. A represents the phases of the cell cycle in different colours. G1 phase refers to the first stage in the interphase, when cells grow and increase the amount of cellular contents (except for DNA) in preparation for cellular division. $S$ phase refers to the moment when DNA is duplicated. G2 phase occurs after the DNA synthesis and constitutes a checking point for DNA damage that would be repaired in this phase prior entering mitosis. In this phase, telomeres might be elongated by telomerase or processed by exonucleases (C-strand resection). M phase refers to mitosis. G0 phase is a quiescence stage in which cells do not divide. Cells only enter G1 phase when receiving the signals to start cell division. B is a schematic representation of the effect that the high percentage of cell death in LoVo might have in the estimation of the population doublings compared to SW480. Blue cells represent viable cells and salmon ones, non-viable ones. A $50 \%$ viability is represented for LoVo cell line to show that the viable cells would have undergone an extra population doubling to reach the same final number of cells. C is a graph relating the cell cycles with the amount of DNA (colour code is same as in figure A). 1 refers to a diploid karyotype and 2, a duplicated diploid karyotype before cell division. $\mathbf{D}$ is an example for two FAC profiles. The one on the left, from SW480, and the one on the right from LoVo. Each phase of the cell cycle is indicated at the bottom with a coloured arrow (same colour code as A and C). The sub-G1 refers to cell death and is indicated with a skull symbol.
each measurement. Figure 5.5 shows a schematic representation of the cell cycle and the amount of DNA for each phase, relating it with a FAC profile for each cell line.

The proliferation index (PI) was calculated using the equation 5.1 described in Liu et al. (2014) to confirm whether the reduction in growth rates observed for SW480 as a result of cell counts correlated with the proliferation index calculated based on cell cycle and to understand the lack of effect on the growth rate for LoVo. In equation 5.1, "S" refers to the percentage of cells in the DNA synthesis phase with $>\mathrm{N}$ and $2 \mathrm{~N}<$ DNA content; "G2/M" to cellular growth and mitosis with 2N DNA content and "G1" to stage after cell division with N DNA content.

$$
\begin{equation*}
\mathrm{PI}=\frac{\mathrm{S}+(\mathrm{G} 2 / \mathrm{M})}{\mathrm{G} 1+\mathrm{S}+(\mathrm{G} 2 / \mathrm{M})} \times 100 \tag{5.1}
\end{equation*}
$$

When performing FAC analysis, the percentage of cell death was assayed as sub-G1 peak (Death) that includes fragmented DNA arising as a consequence of apoptosis and/or necrosis. Therefore, the sub-G1 peak was considered as measurement of total death. I confirmed our previous results using Trypan blue for cell viability and found that control LoVo cells had a significantly ( $\mathrm{p}<0.0001$ ) greater percentage of cell death compared to SW480. Accordingly, figure 5.6.A shows that in control cells, the proliferation index (PI) was usually greater (except for passages 7 and 8) in LoVo than in SW480 as cells needed to divide more actively to counteract the effect of the high death rate. However, under BIBR1532 treatment, the PI was very similar in both cell lines, although when compared with non treated cells, SW480 had greater PI under the treatment, what might reflect a counteracting mechanism against the slower growth rate observed in figure 5.4.

Furthermore, in primary cell lines that reach senescence after a certain number of replication cycles, the cell division rate measured as population doubling per day (PD/day) decreases with cumulative passages (as shown in figure 3.6.B). On the contrary, in cancer cells that can divide indefinitely this rate is independent on the number of passages and it can be very variable.

Figure 5.6.B compares the cell division rate between both cells under each treatment. A large range (from 0.09 to $0.97 \mathrm{PD} /$ days) was observed for different passages, but in control SW480 cells the rate seemed to be pretty constant, around $0.6 \mathrm{PD} /$ day from passages 5 to 13 , while under treatment it was more variable, usually with a lower value, confirming the results obtained with the growth curves. Interestingly, the cell division rate in control LoVo cells increased over $0.8 \mathrm{PD} /$ day usually after two passages with a lower rate ( $\approx 0.4$ PD/day) but in BIBR1532 treated cells, this increase occurred in alternate passages after a very low ( $<0.3 \mathrm{PD} /$ day) cell division rate, what might be related with the higher percentage of cell death in passages $9,11,13,15$ and 17 shown in figure 5.7 that summarises the cell cycle distribution for each cell line under both conditions (control and drug-treated) for all the passages analysed.

The most relevant feature from the FAC profiles (figure 5.7) was the significant increase ( $p<0.0001$ estimated by paired t-test) in the proportion of S phase for SW480 treated cells compared to controls suggesting that cell divisions might have been promoted in this cell line upon the treatment to counteract the potential cytostatic effect of the drug. However, a slight increase in cell death for control SW480 cells that was not apparent in BIB1532 treated cells was also observed, thus having no relation with the treatment. Furthermore, as LoVo cell growth seemed not to be affected by the treatment, no significant changes were observed among the cell cycle, and due to the variability of the FAC data, no conclusions were drawn.


Figure 5.6: Proliferation index and cell division rate for cancer cells treated with BIBR1532. A shows the proliferation index (PI) in percentage, calculated as described in equation 5.1. B shows the cell division rate in number of population doubling (PD) per day for each passage. In both $\mathbf{A}$ and $\mathbf{B}$ graphs, SW 480 is represented in orange colour and LoVo in purple. Graphs on the left refer to control and graphs on the right to treated cells.


Figure 5.7: Cell cycle analysis by FACs. The graphs on the left represent percentages of cells in each phase of the cell cycle for control cells (SW480 at the top and LoVo at the bottom). The graphs on the right represent percentages of cells in each phase of the cell cycle for each counterpart on the left but BIBR1532 treated. Blue colour refers to cell death measured by sub G1; red to G1 with DNA content $=\mathrm{N}$; purple to $\mathrm{G} 2 / \mathrm{M}$ with DNA content $=2 \mathrm{~N}$ and green to $S$ phase with DNA content $>\mathrm{N}$ and $2 \mathrm{~N}<$.

### 5.3.3 Telomere dynamics and telomerase activity

To study the relative percentage of telomerase inhibition reached after long term treatment with $30 \mu \mathrm{M}$ of BIBR1532, the relative telomerase activity was measured by TRAP assay at passage 13 for SW480 and LoVo for both, control and treated cells. This corresponds to PD 27.4 and 24.9 for SW480 (control and treated respectively) and PD 26.5 and 25.8 for LoVo (control and treated respectively). Graph on figure 5.8.B indicates that there was not a significant reduction in telomerase activity for any of the treated cell lines compared to their respective controls, suggesting, a) inefficient drug delivery, b) telomerase activity reactivation after a prolonged inhibition period (Delhommeau et al., 2002) or c) underestimation due to the semi-quantitative nature of the method for measuring telomerase activity. However, telomerase has been proposed to have other functions than elongating telomeres (reviewed in Martínez \& Blasco (2011) and Chiodi \& Mondello (2012)) and as telomere shortening usually occurs in colonic tumour cells that are telomerase positive (Takagi et al., 2000), telomere shortening rates were studied to evaluate if a direct effect on telomere length was observed after long-term treatment with BIBR1532.

To study telomere length dynamics, two telomeres, $12 q$ and XpYp were measured by STELA at two time points, from beginning to end of the cell culture ( 3.5 to 31.2 and 4.0 to 30.6 PDs for SW480 and LoVo respectively) for controls and ( 2.3 to 28.3 and 3.6 to 29.7 PDs for SW480 and LoVo respectively) for BIBR1532 treated cells. Figure 5.9 shows that both colorectal cancer cell lines had very short $12 q$ and XpYp telomeres compared to other cancer cell lines (data not shown) agreeing with previous studies reviewed in Bertorelle et al. (2014). In addition, even with the presence of telomerase activity shown on figure 5.8, telomeres shortened in SW480 comparing start and end time points for both conditions as previously described by (Takagi et al., 2000) and the telomere shortening rate (calculated using equation 1 in chapter 3) was significantly greater in treated than in control cells (Fisher's test).

Surprisingly, telomere length seemed to be more strongly maintained by telomerase in the LoVo cell line as even in the control, telomeres did not shorten (as previously described in Kuranaga et al. (2001) after long-term
treatment with anticancer drugs). The XpYp telomere was significantly shorter at the end of the cell culture in treated than control cells for SW480 but not for LoVo (despite of the cumulative PDs reached were lower for treated than controls). Conversely, the $12 q$ telomere was not different at the end of the cell culture for any cell line independently on the condition, suggesting that the XpYp might be more strongly maintained than $12 q$ telomere (see table 5.2). Interestingly, the telomere elongation in LoVo seemed to be reduced in treated cells compared to controls for the XpYp telomere (from 42.3 to $18.4 \mathrm{bp} / \mathrm{PD}$ ) and not for 12 q , suggesting that despite the absence of telomere shortening, the elongation process in LoVo was somehow interrupted more strongly for XpYp .

A


B


Figure 5.8: TRAP assay on cell pellets from cancer cells treated with $\mathbf{3 0} \boldsymbol{\mu M}$ of BIBR1532. A is an example for a TRAP assay gel. The cell lysate loaded to each lane is indicated at the top. hi: heat inactivated. B is a bar graph indicating the relative telomerase activity as a measurement of the relative intensity of the first PCR band ( $1^{\text {st }} \mathrm{PCR}$ ) to the internal control (IC) on figure A. Means and SE are shown for two replicate experiments.

A


B


Figure 5.9: Scatter plots for telomere length assayed by STELA. A shows XpYp (left) and 12q (right) median telomere lengths for SW480 at two time points. B shows XpYp (left) and 12q (right) median telomere lengths for LoVo at two time points. In all graphs, yellow colour indicates control ( $0.3 \% \mathrm{DMSO}$ ) and red colour indicates cells treated with $30 \mu \mathrm{M}$ of BIBR1532. Medians and interquartile ranges are shown. Interquartile ranges and P values from Exact Fisher's test are shown.

|  |  | XpYp |  | Diff | 12q |  | Diff |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Control | BIBR1532 |  | Control | BIBR1532 |  |
| SW480 | PD | 27.7 | 26.0 |  | 27.7 | 26.0 |  |
|  | mTL Start (bp) | 1943 | 2226 |  | 523.1 | 1261 |  |
|  | mTL End (bp) | 1881 | 1558 |  | 329.3 | 305.3 |  |
|  | TSR (bp/PD) | 2.2 | 25.7 | -23.5 | 7.0 | 36.8 | -29.8 |
| LoVo | PD | 26.6 | 26.1 |  | 26.6 | 26.1 |  |
|  | mTL Start (bp) | 1799 | 2151 |  | 207.9 | 300.8 |  |
|  | mTL End (bp) | 2924 | 2632 |  | 289.9 | 385.7 |  |
|  | TLR (bp/PD) | 42.3 | 18.4 | -23.9 | 3.1 | 3.3 | -0.2 |

Table 5.2: Telomere lengths and telomere shortening rates for SW480 and LoVo cell lines.
PDs: Population Doublings; bp: base pairs; mTL: median telomere length; TSR: telomere shortening rate. TLR: telomere lengthening rate. Diff: difference of TSR or TLR between control and treated cells for each telomere.

### 5.4 Discussion

Telomerase is present in a high proportion of human cancer cells ((Kim et al., 1994) and (Shay \& Bacchetti, 1997)) and thus its activity has been proposed to be related with the acquisition of malignancy during tumorigenesis (Chadeneau et al., 1995) as it is needed for replicative immortality. Accordingly, there are many studies trying to inhibit telomerase using different drugs in an attempt to discover a cancer treatment (reviewed in Rankin et al. (2008) and Phatak \& Burger (2007)).

In this chapter, the telomerase inhibitor BIBR1532 has been used in two different cancer cell lines: LoVo (MMR-deficient, MSH2 ${ }^{-/-}$) and SW480 (MMRproficient, $\mathrm{APC}^{-/-}$). The microsatellite instability (MSI) in LoVo was described to be two fold higher compared to HT1080 (a MMR proficient fibrosarcoma cell line) Hanford et al. (1998) and as explained in chapter 4, the MMR-deficient cell line (LoVo) had a high telomere instability driven by the MMR defect (Mendez-Bermudez \& Royle, 2011). This might impact on telomere shortening and therefore, have an additive effect under telomerase deficiency. Using the BIBR1532 directly onto cell lysates a higher level of telomerase inhibition was reached for SW480 (92.6\%) than for LoVo (58.5\%); however, when the drug was used in cell culture for $\approx 60$ days, no telomerase downregulation was measured using the semiquantitative method TRAP assay. Interestingly, cell growth curves indicated a slower cell growth for SW480 treated cells compared to controls that correlated with an increase in $S$ phase arrest for treated cells shown by flow cytometry analysis (FAC) but no effect was shown for the LoVo cell line, suggesting that either this cell line was more resistant to the antiproliferative effect of BIBR1532 than SW480 or that the drug was not efficiently delivered to this cell line.

Subsequently, the proliferation index (PI) based on the FAC profile was estimated and it was found that the MMR-deficient cancer cell had a greater PI than the MMR-proficient one, specially in the controls, what was due to a very high percentage of cell death in LoVo calculated as subG1 peak compared to SW480. Previous experiments performed by Liu et al. (2014) for telomerase
downregulation in the SW480 cancer cell line using an shTERT showed a reduction in cell growth with an increase in the proportion of apoptotic cells (by Annexin V staining, TUNEL assay and transmission electron microscopy) that reduced the tumour formation in vivo. Jiang et al. (2003) also showed an increase in apoptosis after inhibition of telomerase in SW480 using antisense oligodeoxynucleotides (As-ODN). Unfortunately, only a mild reduction in the cell growth that did not occur with an increase of cell death (as sub-G1 peak) was observed, showing that our drug conditions were sub-optimal.

Subsequently, the telomere length was analysed by STELA and the telomere shortening rates (TSR) were estimated. Despite the lack of evidence for telomerase inhibition measured by TRAP assay, it was found that in SW480 treated cells telomeres shortened at a significantly higher rate than in controls ( 23.5 and $29.8 \mathrm{bp} / \mathrm{PD}$ faster for XpYp and $12 q$ respectively), suggesting that the slower cell growth described before might be a consequence of the increased shortening rate. However, this increment in the TSR was not enough to show an effect on cell death. Interestingly, telomeres were more strongly maintained in the MMR-deficient than in the MMR-proficient cell as telomeres in LoVo lengthened during the cell culture even in the absence of the drug. Accordingly, the high percentage of cell death, the very short median telomere lengths and the strong maintenance of them in LoVo, suggests that treatment of this cell line with a suitable telomerase inhibitor might be an interesting approach for tumour inhibition. Furthermore, it was observed that the XpYp telomere lengthened at a slower rate ( $23.9 \mathrm{bp} / \mathrm{PD}$ less) in treated than control LoVo cells but this effect was not observed for 12q, and no effect on cell growth dynamics was observed.

Treatment of the LoVo cell line with anticancer drugs such as CDDP (cisdiamminedichloroplatinum) and 5FU (5-fluorouracil) resulted in an increase in telomerase activity during the first 15 cumulative PDs and a continuous telomere elongation with increasing PDs (up to 39 cumulative PDs); however, the telomere lengthening was different depending on the initial telomere length ( +93 bp with 2.80 kb length and +45 bp with 4.19 kb length) (Kuranaga et al., 2001). In the present experiment, the starting telomere lengths for LoVo
were shorter than for SW480 in controls and treated cells, suggesting that telomere elongation in this cell line might be needed to continue cell proliferation. Interestingly, El-Daly et al. (2005) showed that the gain of telomeric repeats in immortalized hTERT-HK1 cells was substantially reduced with increasing concentrations of BIBR1532, suggesting a similar effect of that found here in LoVo for the XpYp telomere. Furthermore, it was shown that a telomerase-independent telomere maintenance mechanism might occur in MMR deficient cells (Rizki \& Lundblad, 2001) sharing characteristics with the recombination-based ALT mechanism (Bechter et al., 2004); this might explain the continuous telomere length maintenance in controls and treated LoVo cells in comparison with MMR-proficient cells if the BIBR1532 treatment would have reduced the telomere elongation by telomerase. In addition, MMR deficiency is known to promote G2/M cell cycle arrest after UV DNA damage (reviewed in Conde-Pérezprina et al. (2012)), although if not exposed to genotoxic agents, MMR-deficient cancer cells can escape apoptosis and restoration of the MMR defect in those cells increases the apoptosis rate (reviewed in Hassen et al. (2016)). Surprisingly, a significantly higher percentage of control LoVo cells in sub-G1 (that mostly accounted for dead, apoptotic cells) compared to SW480 was found and the percentage of dead cell did not increased upon treatment (possibly due to not efficient drug delivery as no evidences were found for telomerase activity reduction) and cell growth curves were not different from control and treated LoVo cells. Conversely, when telomerase was previously inhibited in the LoVo cell line using a dominant negative (DN-hTERT), cell proliferation was completely stopped (Hahn et al., 1999). The proposed mechanism by which BIBR1532 inhibits telomerase is due to its direct interaction with the enzyme and/or the telomere, directly affecting the enzyme translocation through the telomere or the dissociation between the enzyme and the DNA template (Pascolo et al., 2002); therefore it does not block the catalytic activity of telomerase but the elongation of the DNA template. In addition, BIBR1532 has been shown to interact with the TEL domain in TPP1, competing with telomerase for its binding to the telomere (Nakashima et al., 2013). Hence, very high concentrations of

BIBR1532 might not only affect the enzymatic activity of telomerase but also impact on the binding of other telomeric proteins such as TRF2 (El-Daly et al., 2005). Furthermore, the telomerase inhibition mechanism by the DN-hTERT depends on the removal of the endogenous hTERT from the telomeres and the cytoplasmic degradation via ubiquitination (Nguyen et al., 2009) having a more detrimental effect in cells where the non-canonical functions of telomerase prevail over the elongation process. In fact, in neuroblastoma cells where telomere elongation is not the most important function of telomerase, the use of DN-hTERT does not impact on telomere length but increases the vulnerability to apoptotic drugs (Samy et al., 2012). In this experiment, an increase in S phase for treated vs control was only found in SW480 but not in LoVo cells, while other telomerase inhibitors such as T-oligos have shown this arrest, cell growth inhibition and expression of DNA damage response proteins, with an increase in senescence that was not accompanied by telomere shortening (Wojdyla et al., 2014), suggesting that for LoVo, drug delivery was not as efficient as in other studies. Taking into account the outcome of this experiment, it was concluded that treatment of the LoVo cell line with BIBR1532 was not as efficient as for the MMR-proficient (SW480) cells. The contradictory results shown in this chapter might be due to several aspects that should be modified in future experiments:

- Control: SW480 might not have been the best control for this experiment as, although it is a similar type of cancer cell than LoVo, it contains other mutations (e.g. $A P C^{-/-}$) that might differently impact on cell cycle dynamics. For future experiments inhibiting telomerase in a MMR-deficient colorectal cancer cell line, the best control would be the same cancer cell line with the restored MMR-mutation (e.g. LoVo and LoVo+chromosome2 Watanabe et al. (2000)).
- Drug: The BIBR1532 drug was used, as it is the most specific synthetic telomerase inhibitor impacting exclusively on the elongation function of telomerase and therefore not affecting other non-canonical functions; however, there are two negative aspects of using this drug: a) high concentrations produce cytotoxic effects and b) the concentrations used in
this experiment are not in the nanomolar range used for clinical treatments, thus the clinical significance is reduced. The PNA drug GRN163 is being used in clinical trials to reduce tumour size by inhibiting telomerase activity and as it has been shown that it reduces telomere length, it might be interesting to see its effect on MMR-deficient cells.
- Drug concentration: The concentration of BIBR1532 used in this experiment was based on previous published work in the same cell lines but no titration experiments were performed. The assumption that the purity of the drug was the same between the published data and our batch might be risky and the performance of titration experiments should be the first step for future experiments.
- Cell proliferation: The use of FAC analysis to analyse cell proliferation by propidium iodide staining might be unreasonable, as other more simple proliferation assays (such as MTT or Ki69) could being used. However, the use of FAC might have been complemented with an additional fluorescent antibody marker, for example a telomere-associated protein such as TRF2 or an apoptosis marker such as Annexin V to increase the information and to give more details about the effect of the drug.
- Other analysis: The telomerase assay used here is a semi-quantitative assay and it would be more reliable if used in conjunction with other assays such as RT-PCR. Besides, comparisons of the relative telomerase activity for the same treatment at different time points might show changes over time that are also important to consider. In addition, other complementary assays might be performed for future experiments such as western blot or qPCR to study protein expression changes in cell cyclerelated proteins (such as p21 or p16 among others) between control and treated cells.


## 6 Telomere dynamics in families

### 6.1 Background

Telomere length (TL) is known to shorten with every cell division in normal cells in culture ((Harley et al., 1990), (Allsopp et al., 1992) and reviewed in Blackburn (1991)) and also it has been shown in humans ((Mondello et al., 1999), (Friedrich et al., 2000) and (Chen et al., 2011)), thus limiting the ability for indefinite cell divisions. Telomere length becomes less variable in healthy seniors (Halaschek-Wiener et al., 2008) and the relationship between TL and disease susceptibility has been extensively studied. Diseases like chronic lymphocytic leukaemia (CLL) (Lin et al., 2010) and other types of cancer such as colorectal cancer (CRC) ((Rampazzo et al., 2010) and (Qin et al., 2014)) or breast cancer (Martinez-Delgado et al., 2011) have been associated with short telomeres compared to normal tissue. Furthermore, mutations in the telomerase gene or other related genes are found in dyskeratosis congenita (DC) patients (Vulliamy et al., 2004), showing a positive association between very short telomeres, disease severity and earlier age of onset (Alter et al., 2012).

Telomerase is known to be active during all the oogenesis stages in the human oocyte (Wright et al., 2001) although in lower levels as the maturation takes place (Turner \& Hartshorne, 2013). In the male germline, telomerase activity also decreases during spermatogenesis and it is inversely correlated to TL (Achi et al., 2000). Sperm TL has been shown to increase during male lifetime ((Baird et al., 2006), (De Meyer et al., 2007) and (Aston et al., 2012)). However, TL in oocytes has been found to be significantly longer than in sperm (Turner \& Hartshorne, 2013) as a result of the fewer cell divisions that occur in the oocyte progenitor cells compared to the sperm. In addition, TERRA and telomerase have been found to be localized at telomeres during mammalian gametogenesis (Reig-Viader et al., 2014), suggesting a very complex regulation of the TL during the development and maturation of human
sex cells.
Results shown in Graakjaer et al. (2006) suggested that despite the presence of telomerase activity in the germline, there was an inherited allelespecific telomere length from the parents that resulted in a fix starting point for each individual's TL, suggesting that the telomere length of the offspring is highly determined by the progenitor's TL in the germline. In addition, other authors have supported the idea that telomere length is not entirely reset in the zygote (De Meyer et al., 2007). Studies in mouse models, where Tert ${ }^{+/+}$ mice were generated after 17 generations of $\mathrm{Tert}^{+/-}$crosses, also support the idea that offspring's TL is determined by the parent's TL in the germline (Chiang et al., 2010). Therefore, it is not surprising that offspring's TL has been associated to paternal age at birth ((Unryn et al., 2005), (De Meyer et al., 2007) and (Kimura et al., 2008)). Furthermore, Graakjaer et al. (2004) suggested that the inherited TL pattern at conception is more important for determining TL during adulthood than any other factor occurring later in life (such as environmental or epigenetic changes affecting telomere length).

Lynch syndrome (LS) is an autosomal inherited disorder caused by a germ line mutation in a MMR gene, usually hMSH2 or hMLH1 (Peltomäki et al., 1993). It is characterized by an early-onset of colorectal cancer (CRC), often with a bias for the right colon and the presence of other extra colonic tumours such as urologic, upper gastrointestinal or gynaecological (reviewed in Lynch et al. (2009) and Boland \& Lynch (2013)).

The study of telomeres in LS families follows two main reasons: on the one hand, the microsatellite instability (MSI) that arises as a consequence of impaired MMR, increases the mutation frequency in repetitive DNA sequences including telomeres (Pickett et al., 2004) with a suggested bias towards losses of repeats as shown by an increase in the telomere shortening rate when the MSH2 protein was downregulated in a primary human cell (Mendez-Bermudez \& Royle, 2011) and in chapter 3. On the other hand, short telomeres have been related to disease anticipation in hereditary syndromes characterised by telomerase gene mutations such as dyskeratosis congenita (DC) (Vulliamy et al., 2004) or p53 mutations like Li-Fraumeni syndrome (Ta-
bori et al., 2007), as well as suggested to be factors explaining other diseases such as vascular dementia (von Zglinicki et al., 2000b). In addition several meta-analysis have shown a positive relationship between short telomeres and cancer risk ((Ma et al., 2011) and (Zhu et al., 2016)) including colorectal cancer Qin et al. (2014) or hereditary breast cancer ((Martinez-Delgado et al., 2011) and (Duggan et al., 2014)). However, no mechanism has been described yet and controversy still remains towards the existence of genetic anticipation in LS. While some studies showed evidence of an earlier onset of cancer in successive generations ((Westphalen et al., 2005), (Nilbert et al., 2009) and (Timshel et al., 2009)), other studies suggested there might be an ascertainment bias generating artificial anticipation (Voskuil et al., 1997), (Tsai et al., 1997), (Stupart et al., 2013) and (Stupart et al., 2014). Interestingly, Stella et al. (2007) found the first evidence of three large deletions in MSH2 having an effect on the age onset of the disease; in particular, the 23 parent/children pair analysed showed a median anticipation of 12 years with no evidence for a birth cohort effect, and all were mutation carriers for any of three large deletions in MSH2, suggesting that genetic anticipation might occur in families carrying only specific germline mutations.

In a pilot study carried out by Dr. Frances Tippins (Univ. of Leicester iB.Sc. student 2013-2014), saliva samples from two LS cohorts were recruited. Results from some of the families analysed showed a lack of significance between telomere length and age for 12 q but not XpYp, although only 39/91 ( $43 \%$ ) samples were analysed. Interestingly, parent/children pair comparisons for the XpYp telomere showed a trend for longer telomeres in wild type children compared to their $\mathrm{MMR}^{+/-}$parents although the age-adjusted telomere length ( $\Delta \mathrm{Tel}$ ) was calculated based on the same cohort using 15 wild type samples (including children). Due to the lack of a control cohort establishing a standard curve for telomere attrition with age, no further conclusions were drawn.

### 6.2 Aims

To understand whether the inheritance of a germline MMR gene mutation in Lynch syndrome (LS) families had an impact on telomere length (TL), and if so, whether a mechanism based on TL might be hypothesised to explain the age of onset of cancer. This aim involved the following objectives:

- To recruit LS and control families with the desired parent/children pair combinations for the TL inheritance study.
- To use single telomere length analysis (STELA) to measure TL in DNA from saliva.
- To study TL in control and LS families to compare TL inheritance between MMR proficient and deficient individuals.
- To study TL inheritance in LS by parent/children pair comparisons to try testing an hypothesis for a mechanism describing genetic anticipation.


### 6.3 Results

### 6.3.1 Recruitment and samples description

Saliva samples from three cohorts were collected, two of which included Lynch syndrome families belonging to either Leicester Royal Infirmary (LRI) or Birmingham Women's Hospital (BWH) and a third one including control families mostly from the Genetics Department in the University of Leicester. The sample collection for the LS families was set up by Dr. Frances Tippins and Dr. Nicola Royle in conjunction with Professor Eamonn Maher and Jonathan Hoffman (BWH) and Dr. Julian Barwell (LRI) as described in section 2.1. Table 6.1 shows a summary of the samples and their main features in terms of sex, age distribution and genetic characteristics. For a more detailed information about the mutations that each family carries, refer to table 6.2. The breakdown of samples collected by Dr. Frances Tippins (FT) or Carmen Garrido (CG) is shown in appendix F.

Figure 6.1 shows that most of the DNA extracted from the saliva samples was high molecular weight (HMW), and therefore, suitable for STELA. Nevertheless, there were two control families (TEL F5 and TEL F10) marked with a red asterisk in figure 6.1.A where the DNA quality was not good enough for STELA and were excluded from the study. In addition, two individuals (TELF6G1P2 and TELF7G2P2) had very degraded DNA and did not generated STELA products, so they were not included in the study. The Oragene ${ }^{\text {TM }}$ DNA saliva kit contains a DNA stabilising agent to control bacterial growth and maintain DNA integrity (Theall et al., 2013). We used the same reagents and procedure to extract HMW DNA as for the samples that gave degraded DNA, so it was hypothesised that in the samples we had to exclude, there was some source of DNA degradation at sample donation, not during DNA extraction.

Saliva samples were grouped into four main groups: LS MMR ${ }^{+/-}$cancerfree, LS MMR ${ }^{+/-}$cancer-affected, LS MMR ${ }^{+/+}$cancer-free and controls (figure 6.2.A). There were 5 samples that did not fit into any of these four groups: 2 of


B


Figure 6.1: DNA quality for control families. A shows a $1 \%$ agarose gel electrophoresis for DNA extracted from saliva in the control families. Red asterisks refer to families that were excluded as no individual generated any STELA products. Green asterisks indicate two additional individuals that were excluded because they did not produce STELA products either. B shows two $1 \%$ agarose gel electrophoresis for DNA extracted from saliva in the two LS cohort. All (91) individuals were included in the study. Codes for the samples are: TEL, control families; LRI, Leicester Royal Infirmary and BWH, Birmingham Women's Hospital. The "F" before a number indicates the family's number. " $G$ " indicates generation and " $P$ " indicates position on the pedigree. $\Lambda$-HindIII was the size marker used at two different concentrations: 10 and $20 \mathrm{ng} / \mu \mathrm{l}$. "kb" indicates kilobases.
them were cancer-affected LS MMR ${ }^{+/+}$patients (thyroid cancer and basal cell carcinoma) and 3 were controls with cancer (acute myeloid leukaemia, breast and testicular cancers). Nevertheless, none of the 5 samples had any LSrelated tumour so they were considered as sporadic tumours, and therefore they were treated as controls. Figure 6.2.B shows the mean age at which the samples were collected for each group. Kruskal-Wallis ranking tests for the age distributions showed significant ( $p=0.007$ ) differences in the age at sampling between cancer-affected and cancer-free LS MMR ${ }^{+/-}$individuals. This is due to the fact that $\mathrm{MMR}^{+/-}$in the second generation (children) have not developed cancer yet.

For all families, a pedigree indicating: code, age, mutation status and cancer information is plotted in appendix I for controls and J for LS, together with a scatter plot showing the median telomere length (mTL) for XpYp and $12 q$ telomeres and the interquartile (IQ) range.


Figure 6.2: Cohort description. A is a pie graph that summarises the number of samples within each group. $\mathbf{B}$ is a scatter plot with bar indicating the mean and standard deviation of the age at which samples were taken for each group. ${ }^{* * *} \mathrm{p}=0.007$.

| N | Lynch syndrome families ( $\mathrm{n}=24$ ) |  |  |  | Control families ( $\mathrm{n}=10$ ) |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | MMR ${ }^{+/-}$( $\mathbf{N}=\mathbf{6 0}$ ) |  | MMR ${ }^{+/+}(\mathrm{N}=31)$ |  | Total 91 | Unrelated controls ( $\mathrm{N}=37$ ) |  | Total <br> 37 |
|  | Cancer | Cancer-free | Cancer | Cancer-free |  | Non LS cancer | Cancer-free |  |
|  | 32 | 28 | 2 | 29 |  | 3 | 34 |  |
| Parent: N (\%) [P/C pairs] | 24 (75) [38] | 3* (10.7) [7*] | 2 (100) [3] | 14 (48.3) [23] | 43 (47.3) [72] | 3 (100) [5] | 15 (44.1) [29] | 18 (48.7) [34] |
| Children: $\mathbf{N}$ (\%) [P/C pairs] | 8 (25) [12] | 25 (89.3) [40] | 0 (0) [0] | 15 (51.7) [19] | 48 (52.7) [72] | 0 (0) [0] | 19 (55.9) [29] | 19 (51.4) [29] |
| MSH2 ${ }^{+-}$: N (\%) | 15 (46.9) | 18 (64.3) | - | - | 33 (36.3) | - | - | - |
| MLH1+ ${ }^{+-} \mathbf{~} \mathbf{N}$ (\%) | 17 (53.1) | 10 (35.7) | - | - | 27 (29.7) | - | - | - |
| $\mathbf{M M R}^{+/+}: \mathbf{N}$ (\%) | - | - | 2 (100) | 29 (100) | 31 (34.1) | 3 (100) | 34 (100) | 37 (100) |
| Sex: N (\%) | F: 14 (43.8) | F: 17 (60.7) | F: 2 (100) | F: 16 (55.2) | F: 49 (53.8) | F: 2 (66.7) | F: 21 (62.2) | 23 (62.2) |
|  | M: 18 (56.2) | M: 11 (39.3) | M: 0 (0) | M: 13 (44.8) | M: 42 (46.2) | M: 1 (33.3) | M: 13 (37.8) | 14 (37.8) |
| Mean age at sample $\pm$ SD | $53.7 \pm 13.5$ | $36.4 \pm 12.4$ | $65.3 \pm 10.9$ | $48.5 \pm 15.9$ | $47.0 \pm 15.8$ | $55.5 \pm 2.8$ | $44.4 \pm 19.0$ | $45.3 \pm 18.4$ |
| Mean age at sample $\pm$ SD parents | $60.0 \pm 7.7$ | $66.3 \pm 2.0$ | $65.3 \pm 10.9$ | $62.7 \pm 7.3$ | $61.6 \pm 7.5$ | $55.5 \pm 2.8$ | $62.1 \pm 10.4$ | $60.9 \pm 9.8$ |
| Mean age at sample $\pm$ SD children | $34.8 \pm 8.2$ | $32.9 \pm 7.0$ | - | $35.1 \pm 8.0$ | $33.9 \pm 7.4$ | - | $30.5 \pm 10.6$ | $\mathbf{3 0 . 5} \pm \mathbf{1 0 . 6}$ |
| Mean age 1st cancer $\pm$ SD | $39.4 \pm 13.4$ | - | $58.5 \pm 9.3$ | - | $40.5 \pm 14.2$ | $43.7 \pm 7.6$ | - | $43.7 \pm 7.6$ |
| Mean age 1st cancer $\pm$ SD parents | $44.5 \pm 11.4$ | - | $58.5 \pm 9.3$ | - | - | $43.7 \pm 7.6$ | - | - |
| Mean age 1st cancer $\pm$ SD children | $23.9 \pm 7.1$ | - | - | - | - | - | - | - |

Table 6.1: Features of the studied groups. The two Lynch syndrome (LS) cohorts from BWH and LRI are grouped in the LS families group. n: number of families; N: number of subjects percentage to the total of each group; P/C: parent-children; SD: standard deviation; F: female; M: male. *Indicates that the number is effectively +1 (individual/pair) as sample BWH F105 G4P9 was treated as children and as parent (see pedigrees in appendix I for control and J for LS families for more information)

### 6.3.2 Telomere length measurement and reproducibility

All saliva samples had a unique code for identification; however, two different types of codes were used per sample. On the one hand, for sample recruitment and DNA extraction, the unique identifier used per individual contained a prefix with the family code (either BWH or LRI for LS families or TEL for control families), that was followed by a number referring to the family number. Next, a "G" indicated the generation and was followed by a number referring to the generation the individual had on the pedigree and finally, a number after a "P" indicated the position in the pedigree. For example, an individual coded as: BWH F105 G4P2, belonged to a LS family recruited in Birmingham Women's Hospital, was included in the pedigree for family 105, referred to the $4^{\text {th }}$ generation and occupied the position 2 in the pedigree. On the other hand, for telomere length analysis a blind code was generated to avoid bias in the measurements based on the known age of the individual. This code contained the first part of the unique code, indicating cohort (BWH, LRI or TEL) and family number, but the "G-" and "P-" were substituted by a letter in alphabetical order depending on the number of individuals per family.

Reproducibility of the method was first assayed by comparing telomere length measurements for the same coded family (BWH F100) between the two students (FT and CG); the letters A, B and C in figure 6.3 anonymised each of the three members for this family. No significant differences were obtained when comparing data generated by either student, confirming that our results were highly reproducible and comparable.

In addition, the same control DNA, extracted from the lymphoblastoid cell line KK, was used in each STELA gel amplified with the flanking primer XpYpE2 (see table in appendix B for primer sequence). Analysis of this control sample allowed the data from saliva samples in different gels to be combined for the same individual and to compare between gels and families. For each sample, at least a duplicate experiment was performed and when possible, members from the same family were analysed in the same Southern blot. Kruskal-Wallis tests, corrected for multiple comparisons, were performed on all the control DNA experiments before pooling data. Only two replicate gels

| Family | Type of mutation/location | C.Significance | Link |
| :---: | :---: | :---: | :---: |
| LRI-F1 | Missense mutation (c.1447C $>$ T; p.Glu483X) in exon 9 of the $h M S H 2$ gene | Risk factor | rs587777421 |
| LRI-F4 | Deletion of exons 9 to 16 of the hMSH2 gene | Pathogenic | DOI 10.1002/humu. 10291 |
| LRI-F5 | Missense mutation (c.1861C>T; p.Arg621X) in exon 12 of the $h M S H 2$ gene | Pathogenic | rs63750508 |
| LRI-F7 | Missense mutation (c.1373T $>\mathrm{G}$; p.Leu458X) in exon 8 of the $h M S H 2$ gene | Pathogenic | rs63750521 |
| LRI-F14 | Deletion of exon 1 of the $h \mathrm{MSH} 2$ gene | Pathogenic | OMIM: 609309.0015 |
| BWH-F5a | Missense mutation (c.2422G>T; pGlu808X) in exon 14 of the $h M S H 2$ gene | Pathogenic | rs34986638 |
| BWH-F11a | Frameshift mutation (c.1035-1036delGAinsT; p.Trp345Cysfs $\times 12$ ) in exon 6 of the $h M S H 2$ gene | Untested | - |
| BWH-F55 | Nonsense mutation (c.2126T $>$ A; p.L709X) in exon 13 of the $h M S H 2$ gene | Untested | - |
| BWH-F71 | Deletion of exon 1 of the hMSH 2 gene | Pathogenic | OMIM: 609309.0015 |
| BWH-F100 | Splice-site mutation (c.2211-1G>C) in exon 13 of the $h M S H 2$ gene | Untested | - |
| BWH-F105 | Missense mutation (c.1030C $>$ T; p.Gln344X) in exon 6 of the $h M S H 2$ gene | Pathogenic | rs63750245 |
| BWH-F108 | Frameshift mutation (c.2562delT; p.Gln855SerfsX37) in exon 15 of the $h M S H 2$ gene | Untested | - |
| BWH-125 | Frameshift mutation (c.2332delT; p.Cys778AlafsX34) in exon 14 of the $h M S H 2$ gene | Untested | - |
| LRI-F12 | Deletion of the entire coding region of the $h M L H 1$ gene | Untested | - |
| BWH-F1a | Deletion of exons 1 to 19 in the $h M L H 1$ gene | Untested | - |
| BWH-F3a | Missense mutation (c. $200 \mathrm{G}>\mathrm{A}$; p.Gly67Glu) in exon 2 of the $h M L H 1$ gene | Pathogenic | rs63749939 |
| BWH-F6a | Deletion (c.1852-1854delAAG; p.Lys618del) in exon 16 of the hMLH1 gene | Pathogenic | rs63751247 |
| BWH-F2 | Framesift mutation (c.1348dupG; p.Asp450GlyfsX29) in exon 12 of the $h M L H 1$ gene | Pathogenic | rs587778906 |
| BWH-F42 | Deletion of exon 6 in the $h M L H 1$ gene | Untested | - |
| BWH-F102 | Missense mutation (c.380G>A; p.Arg127Lys) in exon 4 of the $h M L H 1$ gene | Likely pathogenic | rs63751595 |
| BWH-F106 | Splice-site mutation (c.1668-1G $>$ A) in intron 15 of the $h M L H 1$ gene | Likely pathogenic | rs267607845 |
| BWH-F107 | Missense mutation (c.199G>A; p.Gly67Arg) in exon 2 of the $h M L H 1$ gene | Pathogenic | rs63750206 |
| BWH-F126 | Frameshift mutation (c.94delA; p.Ile32SerfsX4) in exon 1 of the hMLH1 gene | Untested | - |
| BWH-F130 | Missense mutation (c.306G>T; p.Glu102Asp) in exon 3 of the $h M L H 1$ gene | Likely pathogenic | rs63751665 |

Table 6.2: Mutation details per family. The top part of the table shows the specific germline mutation that each family carries for the human MSH2 gene ( $h M S H 2$ ). The bottom part of the table shows the specific germline mutation that each family carries for the human MLH1 gene ( $h M L H 1$ ). For all mutations: "c." indicates location in the coding sequence; "p." location in the protein sequence; ">" refers to a point mutation; "-" between two numbers indicates frameshift mutation; "X" indicates stop codon; "del" indicates deletion and "ins" indicates insertion.
out of 128 gels were significantly different from the rest (representing a $1.6 \%$ error in the method) and therefore, data from those replicate gels were excluded from further analysis (see appendix G and H). Figure 6.5 shows four STELA gels from two different LS families and the quantification, as an example.


Figure 6.3: Technical reproducibility between students. The scatter plots show median telomere lengths (mTL) and interquartile ranges for XpYp (left) and 12 q (right) telomeres in family BWH 100. Individuals were coded as $\mathrm{A}, \mathrm{B}$ and C to avoid bias in the interpretation. Data in purple was generated by Frances Tippins and data in green generated by Carmen Garrido. Medians and interquartile ranges are shown.

Telomere length for XpYp and 12 q telomeres were measured for all 128 individuals and to decide whether means or medians would describe our samples better, D'Agostino \& Pearson omnibus normality test was run for each individual and telomere. Surprisingly, the percentage of individuals with normally distributed telomeres was greater for control individuals than LS for XpYp ( $40 \%-21 / 52-$ vs $33 \%-25 / 76$-) and $12 q$ ( $56 \%-29 / 52-$ vs $41 \%-31 / 76-$ ). Nevertheless, there was no difference between parent/children percentages in normality between both cohorts, suggesting that the telomere length distribution per individual was not affected by age. Interestingly, when identifying outliers using the ROUT method ( $\mathrm{Q}=1 \%$ ), a greater percentage of LS individuals with outliers for $\mathrm{XpYp}(19 \%-14 / 76-v s 16 \%-6 / 52-$ ) and especially for 12 q ( $33 \%-25 / 76-$ vs $17 \%-9 / 52$-) was observed. This might explain why the percentage of normality decreased in the LS cohort compared to controls but it also suggested that the presence of a MMR deficiency might have an impact on telomere length dynamics. In addition, correlation analysis between XpYp and 12 q telomere lengths for each individual showed that both telomeres were significantly correlated ( $p<0.0001$ and $p=0.002$ for controls and LS respectively). However, the r values of the linear regressions were smaller in LS individuals ( $\mathrm{r}=0.41$ ) than in controls ( $\mathrm{r}=0.63$ ), suggesting a greater TL heterogeneity in the LS cohort (figure 6.4), although the $95 \%$ confidence intervals overlapped somehow ( 0.20 to 0.59 for LS and 0.42 to 0.77 for controls).


Figure 6.4: Correlation of XpYp and 12 telomere lengths. Red colour indicates 52 control individuals ( 37 from the control cohort and $15 \mathrm{LS} \mathrm{MMR}^{+/+}$parents) and purple, 76 LS individuals ( $\mathrm{MMR}^{+/-}$or $\mathrm{MMR}^{+/+}$children).






 The unique code for each individual is shown at the bottom of the Southern blot and in the X axis of the scatter plots; numbers in brackets indicate the age (in years) at sample donation.

### 6.3.3 Telomere shortening with age

To analyse the relationship between the age (at sample donation) and the telomere length (TL) in normal families, 10 healthy families were mostly recruited from the Genetics Department in the University of Leicester, comprising 37 individuals with a mean age of 45.3 ( $\pm 18.4 \mathrm{SD}$ ) years. In addition, 15 samples from the LS cohort that were $\mathrm{MMR}^{+/+}$and spouses of the $\mathrm{MMR}^{+/-}$ parents (therefore, from LS-non related families) were included as controls to increase the sample size of the control cohort. In total, the control cohort contained 52 samples that were used to draw a standard curve for the telomere attrition rate with age for subsequent comparisons with the samples from the LS families.

Two telomeres, $12 q$ and XpYp were measured by STELA as described in section 2.2.2.8 and telomere length (TL) data was plotted against age for all the individuals. The linear regression was not strong ( $\mathrm{r}<-0.3$ ) and not significant ( $p=0.0642$ for XpYp and $p=0.0426$ for $12 q$ ), suggesting weak relationship between TL and age (figure 6.6.A). As discussed in section 1.2.1.3 (view table 1.4), the main disadvantage of STELA is the underestimation of TL due to the inability of measuring very long telomeres. This is, to our knowledge, the first time that STELA has been used to measure individual's TL from DNA extracted from saliva in a cohort with such a large age-distribution cohort ( 18.2 to 82.6 years). To address this issue, we reduced the number of STELA products per track up to a maximum of 10 , and therefore it was assumed that the lack of significance between age and TL was due to underestimation of the TL in younger samples and not to a bias for shorter telomeres. Thus, we performed a piecewise regression as described in Wagner et al. (2002). This is a segmented linear regression in which the dependent variable (TL) has different linear relationships with the independent variable (age) for at least two different intervals. We used the $25 \%$ percentile for age ( 35 years) as a breaking point and two linear regression curves were generated per telomere, one comprising the age interval 18 to 35 years and another for individuals older than 35 years. As shown in figure 6.6.B, the two age intervals had different slopes and samples over 35 years had a regression analysis
slope that was significantly different from zero (with $\mathrm{r}=-0.3$ and $\mathrm{r}=-0.4$ for XpYp and 12 q respectively) in contrast to samples from individuals below 35 years. Human telomeres shorten with age at different rates during embryogenesis (Cheng et al., 2013) and in childhood (Frenck et al., 1998) but it has been largely agreed that the rate is constant during adulthood (Slagboom et al. (1994) and Chen et al. (2011)), even in different tissues (Daniali et al., 2013). Since regression results for the $75 \%$ of the individuals, including those older than 35 years, showed a significant negative linear relationship between telomere length and age that was not shown for individuals younger than 35 years, we assumed that a certain level of telomere length underestimation occurred in the $25 \%$ of our samples. However, we could not exclude those individuals because parent/children pairs were needed for subsequent analyses, therefore, we assumed that telomere length in younger than 35 years shortened with age at similar rate than in older individuals and we used the regression equations 6.1 and 6.2 obtained for the $75 \%$ of the data (including older than 35 years individuals) to transform the telomere length for the remaining $25 \%$ of the samples.

$$
\begin{equation*}
X p Y p \text { expected } T L=(-35.12 \times \text { Age })+6870 \tag{6.1}
\end{equation*}
$$

$$
\begin{equation*}
12 q \text { expected } T L=(-31.14 \times \text { Age })+6905 \tag{6.2}
\end{equation*}
$$

The mean expected TL ( $\mathrm{mTL}_{\text {exp }}$ ) for all individuals with $<35$ years was 6017 and 6149 bp for XpYp and 12q respectively, nevertheless, the observed TL mean ( $\mathrm{mTL}_{\text {obs }}$ ) was 4990 and 5202 bp for XpYp and $12 q$ respectively. Using the observed and the expected values, we estimated that the telomere length for this group of individuals had been underestimated by $17.1 \%$ and $15.4 \%$ for XpYp and 12 q respectively using the following equation:

$$
\begin{equation*}
\% \text { of underestimation }=100-\frac{100 \times \mathrm{mTL}_{\mathrm{obs}}}{\mathrm{mTL}_{\exp }} \tag{6.3}
\end{equation*}
$$

Therefore, we transformed the observed telomere length for each individual younger than 35 years multiplying by 1.171 and 1.154 for $X p Y p$ and $12 q$ respectively. When the transformed data (for younger than 35 years) and the raw data (for older than 35 years) for telomere length was plotted against age, there was a very significant ( $\mathrm{p}=0.0002$ for XpYp and $\mathrm{p}<0.0001$ for 12q) association for both telomeres with age, and the linear regression model significantly improved ( $\mathrm{r}=-0.5$ for XpYp and $\mathrm{r}=-0.6$ for 12 q ) and allowed to include all 52 individuals belonging to the control cohort as shown in figure 6.6.C.

To study telomere shortening with age in the Lynch syndrome cohort only data for $\mathrm{MMR}^{+/-}$(parents and children) and $\mathrm{MMR}^{+/+}$children was included, as the $\mathrm{MMR}^{+/+}$parents were non-LS related and were included in the control cohort. Regression analysis for both telomeres showed a slightly significant negative relationship between age and median telomere length (figure 6.7.A. However, as the age range of this cohort was not significantly different ( $\mathrm{p}=$ 0.4237 for a rank Mann Whitney test) to the age range in the control cohort, we expected that telomere length might have been underestimated due to our method for individuals younger than 35 years. Thus, a piecewise regression using the same breaking point (35 years) as in the control cohort was also performed.

Contrary to what happened in the piecewise regression for the controls, in this cohort the regression line for younger ( $<35$ years) individuals had a positive slope, suggesting that the telomere length for the oldest individuals in this age interval was possibly more accurately measured. This effect might be explained if telomeres in LS individuals were shorter than in controls, and thus for the same age, a smaller degree of underestimation would be made. In addition, for the second age interval ( $>35$ years), the regression lines showed higher $r$ values suggesting a better fit of the model and the significance of the curves was greater (figure 6.7.B). The data was transformed as previously described for the control cohort, but in this case, the equations used to calculate the estimated telomere length for the first age interval ( $<35$ years) were:
A


| $X p Y p$ | $12 q$ |
| :--- | :--- |
| $Y=(-15.94 \times X)+5668$ | $Y=(-13.02 \times X)+5772$ |
| $r=-0.2$ | $r=-0.2$ |
| $r^{2}=0.067$ | $r^{2}=0.079$ |
| $p=0.06$ | $p=0.04$ |

B


| $\mathrm{XpYp>35}$ years | $12 q>35$ years |
| :--- | :--- |
| $\mathrm{Y}=(-35.12 \times X)+6870$ | $\mathrm{Y}=(-31.14 \times \mathrm{X})+6905$ |
| $r=-0.3$ | $r=-0.4$ |
| $r^{2}=0.099$ | $r^{2}=0.149$ |
| $p=0.05$ | $\mathrm{p}=0.02$ |
| $\mathrm{XpYp}<35$ years | $12 q<35$ years |
| $Y=(-78.75 \times X)+6968$ | $Y=(-61.84 \times X)+6753$ |
| $r=-0.5$ | $r=-0.4$ |
| $r^{2}=0.220$ | $r^{2}=0.175$ |
| $p=0.10$ | $p=0.16$ |

C


| $X p Y p$ | $12 q$ |
| :--- | :--- |
| $Y=(-34.31 \times X)+6807$ | $Y=(-31.01 \times X)+6840$ |
| $r=-0.5$ | $r=-0.6$ |
| $r^{2}=0.249$ | $r^{2}=0.308$ |
| $p=0.0002$ |  |
|  |  |
|  |  |

Figure 6.6: Telomere shortening with age in control individuals. For all graphs, blue is XpYp and green, 12q telomeres. $\mathbf{A}$ is a regression analysis of the raw data for control (non-LS related) individuals (52). Table on the right shows curves equations, $r$ and $r^{2}$ values as a measurement of the goodness of fit and $p$ values for the slope confidence. $\mathbf{B}$ is a piecewise regression with breaking point at 35 years for the same data in A . Here, yellow is XpYp for younger than 35 years and brown, 12 q for the same age interval. The table on the right shows same information as in A for all four curves. $\mathbf{C}$ is the final regression curve after data transformation.

$$
\begin{equation*}
X p Y p \text { expected } T L=(-24.50 \times \text { Age })+6247 \tag{6.4}
\end{equation*}
$$

$$
\begin{equation*}
12 q \text { expected } T L=(-21.51 \times \text { Age })+6353 \tag{6.5}
\end{equation*}
$$

In this cohort, the mean expected telomere length ( $\mathrm{mTL}_{\text {exp }}$ ) for the $<35$ years group was 5539 and 5732 bp for XpYp and $12 q$ respectively, while the mean for the observed telomere length ( $\mathrm{mTL}_{\text {obs }}$ ) was: 5259 (for XpYp ) and 5361 (for $12 q$ ) bp. As predicted before when describing the piecewise regression, the underestimation in the telomere length measurements was lower than in the control cohort: $5.06 \%$ and $6.47 \%$ for XpYp and 12 q respectively. Subsequently, data for the observed median telomere length (in the first age interval individuals, $<35$ years) was multiplied by 1.051 and 1.065 for XpYp and 12 q respectively and when transformed data for individuals $<35$ years was plotted together with raw data for the second age interval against age, the model fit improved significantly as shown in figure 6.7.C.

Curve comparisons performing ANCOVA analysis (see table 6.3) between controls and $\mathrm{MMR}^{+/-}$individuals showed no significantly different slopes or Y-intercepts for any telomere as shown in figure 6.8.A (confirming previous results from Seguí et al. (2013)). Nevertheless, telomeres seemed to shorten at a slower rate in $\mathrm{MMR}^{+/-}$compared to controls ( 17 vs $34 \mathrm{bp} /$ year for XpYp and 15 vs $31 \mathrm{bp} /$ year for 12 q ), suggesting they might be shorter over generations in $\mathrm{MMR}^{+/-}$, what might result in an underestimation of the telomere attrition rate in cross-sectional studies as hypothesised by Holohan et al. (2015). Furthermore, the Y-intercepts (predicted telomere length at birth) were smaller in $\mathrm{MMR}^{+/-}$than in controls ( 5824 vs 6807 bp for XpYp and 6011 vs 6840 bp for $12 q$ ), suggesting that the telomere attrition rate found might be explained, at least partially, by the baseline telomere length at birth. Comparisons within the LS cohort between MMR proficient ( $\mathrm{MMR}^{+/+}$) and deficient ( $\mathrm{MMR}^{+/-}$) samples showed that telomeres shortened at a slower (but not significantly different) rate for $\mathrm{MMR}^{+/-}$than $\mathrm{MMR}^{+/+}$(17 vs $30 \mathrm{bp} /$ year for XpYp and 15 vs 27 bp/year for 12 q ) and that $\mathrm{MMR}^{+/-}$had smaller (but not signif-


| XpYp | $12 q$ |
| :--- | :--- |
| $\mathrm{Y}=(-16.63 \times \mathrm{X})+5832$ | $\mathrm{Y}=(-11.40 \times \mathrm{X})+5796$ |
| $r=-0.3$ | $r=-0.2$ |
| $r^{2}=0.079$ | $r^{2}=0.053$ |
| $p=0.01$ | $p=0.05$ |



| $\mathrm{XpYp>35}$ years | $12 q>35$ years |
| :--- | :--- |
| $Y=(-24.50 \times X)+6247$ | $\mathrm{Y}=(-21.51 \times X)+6353$ |
| $r=-0.4$ | $r=-0.4$ |
| $r^{2}=0.128$ | $r^{2}=0.139$ |
| $p=0.01$ | $p=0.01$ |
| $X p Y p<35$ years | $12 q<35$ years |
| $Y=(81.97 \times X)+2988$ | $Y=(63.62 \times X)+3591$ |
| $r=0.4$ | $r=0.3$ |
| $r^{2}=0.135$ | $r^{2}=0.114$ |
| $p=0.05$ | $p=0.08$ |



| $X p Y p$ | $12 q$ |
| :--- | :--- |
| $Y=(-21.80 \times X)+6187$ | $Y=(-19.14 \times X)+6271$ |
| $r=-0.4$ | $r=-0.4$ |
| $r^{2}=0.122$ | $r^{2}=0.129$ |
| $p=0.002$ | $p=0.001$ |

Figure 6.7: Telomere shortening with age in Lynch syndrome patients. For all graphs, blue is XpYp and green, 12 q telomeres. A is a regression analysis of the raw data for LS individuals (76). Table on the right shows curves equations, $r$ and $r^{2}$ values as a measurement of the goodness of fit and $p$ values for the slope confidence. B is a piecewise regression with breaking point at 35 years for the same data in A. Here, yellow is XpYp for younger than 35 years and brown, $12 q$ for the same age interval. The table on the right shows same information as in A for all four curves. $\mathbf{C}$ is the final regression curve after data transformation.
icantly different) Y-intercepts than $\mathrm{MMR}^{+/+}$(5824 vs 6933 bp for XpYp and 6011 vs 6866 bp for 12q) (figure 6.8.B). Interestingly, no significant correlation was found between telomere length and age for $M L H 1^{+/-}$samples (see table 6.3), confirming previous results in Bozzao et al. (2011a). In addition, $\mathrm{MSH2}^{+/-}$samples had the smallest Y-intercept (5688 and 5960 bp for XpYp and 12 q respectively; figure 6.8.C), suggesting that $\mathrm{MSH}^{+/-}$carriers either inherit shorter telomeres than $\mathrm{MSH} 2^{+/+}$or suffer a higher telomere shortening rate during development.


Figure 6.8: Regression curves comparison for telomere length with age. A compares the telomere shortening with age between controls in red ( 52 individuals) and LS in purple ( 76 individuals) for XpYp (left graph) and 12 q (right graph). B compares the telomere length with age between $\mathrm{MMR}^{+/-}$in light green ( 60 individuals) and wild type ( $\mathrm{MMR}^{+/+}$) children with a $\mathrm{MMR}^{+/-}$parent in blue ( 16 individuals). $\mathbf{C}$ compares the telomere length with age between $\mathrm{MLH} 1^{+/-}$in yellow ( 27 individuals) and $\mathrm{MSH} 2^{+/-}$in green (33 individuals).

|  | $\mathbf{X p Y p}$ |  |  |  | 12q |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Slope <br> $\pm$ error | $\begin{gathered} \text { Y- intercept } \\ \pm \text { error } \end{gathered}$ | r | p | Slope $\pm$ error | $\begin{gathered} \text { Y- intercept } \\ \pm \text { error } \end{gathered}$ | r | p |
| Control | $-34 \pm 8$ | $6807 \pm 448$ | -0.5 | 0.0002 | $-31 \pm 7$ | $6840 \pm 349$ | -0.6 | $<0.0001$ |
| MMR ${ }^{+/}$ | $-17 \pm 7$ | $5824 \pm 333$ | -0.3 | 0.0204 | $-15 \pm 6$ | $6011 \pm 302$ | -0.3 | 0.0187 |
| ANCOVA p | 0.1035 | 0.4786 |  |  | 0.0843 | 0.6320 |  |  |
| MMR ${ }^{+/}$ | $-17 \pm 7$ | $5824 \pm 333$ | -0.3 | 0.0204 | $-15 \pm 6$ | $6011 \pm 302$ | -0.3 | 0.0187 |
| MMR ${ }^{+/+}$ | $-30 \pm 22$ | $6933 \pm 875$ | -0.3 | 0.2047 | $-27 \pm 16$ | $6866 \pm 636$ | -0.4 | 0.1179 |
| ANCOVA p | 0.5117 | 0.0192 |  |  | 0.4967 | 0.0693 |  |  |
| MSH2 | $-18 \pm 9$ | $5688 \pm 418$ | -0.4 | 0.0480 | $-17 \pm 8$ | $5960 \pm 382$ | -0.4 | 0.0451 |
| MLH1 | $-15 \pm 11$ | $6023 \pm 516$ | -0.3 | 0.1604 | $-14 \pm 10$ | $6087 \pm 491$ | -0.3 | 0.1872 |
| ANCOVA p | 0.8543 | 0.0337 |  |  | 0.8205 | 0.1828 |  |  |

Table 6.3: Regression curves comparison for telomere length attrition with age in saliva samples. The data refers to the linear regression curves shown in figure 6.8. "ANCOVA $p$ " indicates the $p$ value for the analysis of covariance in linear regression curve comparison. "MMR ${ }^{+/+}$" refers to wild type children belonging to the LS cohorts and thus, having a mutation carrier progenitor." $r$ " shows the goodness of the fit; " $p$ " indicates the significance of the linear regression. Significant values are shown in italics.

It still remains unknown whether there is a sex bias in telomere length and/or shortening with age, as some authors have described women with longer telomeres than males ((Benetos et al., 2001), (Cawthon et al., 2003), (Steenstrup et al., 2013) and (Steenstrup et al., 2013) among others), whereas other authors have found men having longer telomeres than women (De Meyer et al., 2007) and in fact, shortening at a faster rate in males than females ((Vaziri et al., 1993) and (Gardner et al., 2014)). When telomere attrition with age was studied accounting for sex in the control cohort, no significant differences were observed, although Y-intercepts were slightly greater in females ( 7282 and 6915 bp for XpYp and 12q respectively) than in males ( 5956 and 6695 bp for XpYp and $12 q$ respectively). No differences were observed either for the telomere shortening with age, but again, a tendency to greater shortening rate in females ( 43 and $32 \mathrm{bp} /$ year for XpYp and 12 q respectively) than in males ( 20 and $29 \mathrm{bp} /$ year for XpYp and 12 q respectively) was observed (figure 6.9.A). Likewise, no differences were observed in telomere length or attrition rates between males and females for the LS cohort, having a similar shortening for XpYp (22 and 21 bp/year) and 12q (20 and 19 bp/year) (figure 6.9.B).


Figure 6.9: Regression curves comparison between sexes for telomere length with age. A compares the telomere shortening with age between males (blue) and females (pink) in the control cohort. B compares the telomere length with age between males (blue) and females (pink) in the LS cohort.

### 6.3.4 Telomere length inheritance

As telomeres shorten with age, all children might be expected to have longer telomeres than their parents. Therefore, the median telomere length was plotted for each family indicating children with circles and parents with triangles (figure 6.10.B). It was found that $42 \%(8 / 19)$ and $26 \% ~(5 / 19)$, for XpYp and $12 q$ telomeres respectively, of the children had at least one telomere shorter than, at least, one parent (in the control cohort). Nevertheless, this percentage was greater when plotting data before transformation (figure 6.10.A), supporting again the idea of telomere length underestimation in the younger section of our cohort.


Figure 6.10: Median telomere length in control families. A is a scatter plot showing the raw median telomere length (mTL) of each control individual per family (before data transformation). B is a scatter plot showing the mTL for each control individual per family after data transformation of the $<35$ years. For all graphs: blue colour refers to XpYp , green to $12 q$, circles are children, triangles parents, black triangles parents with cancer and red circles around children highlight the ones with shorter mTL than, at least, one parent. See pedigrees in appendix I for further family information.

The percentage of children with shorter telomeres than their parents was also studied in the LS families and compared them with controls. A smaller percentage of children with shorter telomeres than, at least, one parent was found compared to the control families (figure 6.11) for XpYp (36\% -17/47-) and for 12 q ( $23 \%-11 / 47-$ ). Interestingly, of the children with shorter telomeres than at least one parent, all except one (LRI F4 G5P4), were mutation carriers in MSH2 families. However, there were similar number of children with shorter telomeres than, at least one parent, comparing $\mathrm{MMR}^{+/+}$and $\mathrm{MMR}^{+/-}$ children in MLH1 families, suggesting different effects on telomere length depending on the mutated gene. In addition, only one of the four $\mathrm{MSH} 2^{+/-}$ cancer-affected children had both telomeres longer than their parents (although it was a difficult family as the mother carried a BRCA2 mutation and the mutation status for this gene was not known for the daughter). However, in the MLH1 families there were three out of four $\mathrm{MLH} 1^{+/-}$cancer-affected children (BWH F42 G6P1, BWH F126 G4P3 and BWHF1a G4P3) with both telomeres longer than their parents, suggesting there might be a tendency for shorter telomeres in MSH2 and for longer in MLH1 families.

To study the relationship between telomere length and genetic anticipation of the disease, Vulliamy et al. (2004) measured the telomere length by TRF in 8 dyskeratosis congenita (DC) families carrying a mutation in the RNA component of the telomerase ( $h T E R C$ ). After adjusting the telomere length to age, comparisons between parent/children ( $\mathrm{P} / \mathrm{C}$ ) pairs showed that children had shorter age-adjusted telomere lengths ( $\Delta \mathrm{Tel}$ ) than their parents compared to 87 unrelated control pairs, suggesting that the significant decrease in the telomere length over generations might be related to anticipation. Here, we hypothesised that if telomere length had an effect on the age of cancer onset and thus contributing to anticipation, we might observe a similar effect to that previously described in DC patients. Therefore, we calculated the estimated telomere length for all the LS and control individuals, as previously described, using the linear regression equations 6.6 and 6.7 for XpYp and $12 q$ respectively. Subsequently, we estimated the age-adjusted telomere length ( $\Delta \mathrm{Tel}$ ) for each individual using equations 6.8 and 6.9 for XpYp and


Figure 6.11: Median telomere length in Lynch syndrome families. A shows the median telomere length (mTL) in LS families carrying an MSH2 germline mutation. B shows the mTL in LS families carrying an MLH1 germline mutation. For both graphs, left part shows data for the XpYp telomere (blue) and right part, for $12 q$ (green). Graph code: circles are children and triangles parents. Grey colour indicates $\mathrm{MMR}^{+/+}$while blue or green indicate $\mathrm{MMR}^{+/-}$for either XpYp or 12q respectively; yellow indicates a $\mathrm{MMR}^{+/+}$parent that carries a BRCA2 germline mutation. Triangles or circles with a black border indicate cancer-affected individuals. Red circles around some children highlight children with shorter telomeres than, at least, one parent. When several branches can be studied independently within one family (see pedigrees on appendix J for more details), a letter A, B or C next to the family indicates the branch. For branch B in family BWH105, three generations can be studied, therefore: triangle shows first generation, hexagon the second and star the third.)
$12 q$ respectively and finally, the change in telomere length over generations was calculated as in Vulliamy et al. (2004) using equation 6.10.

$$
\begin{align*}
& \mathrm{eTL}_{\mathrm{XpY}}=(-34.31 \times \mathrm{X})+6870  \tag{6.6}\\
& \mathrm{eTL}_{12 \mathrm{q}}=(-31.01 \times \mathrm{X})+6840  \tag{6.7}\\
& \Delta \mathrm{XpYp}=\text { oTL }_{\mathrm{XpYp}}-\mathrm{eTL}_{\mathrm{XpYp}} \tag{6.8}
\end{align*}
$$

$$
\begin{equation*}
\Delta 12 q=o \mathrm{TL}_{12 q}-\mathrm{eTL}_{12 q} \tag{6.9}
\end{equation*}
$$

$$
\begin{equation*}
\text { Change in telomere length }=\Delta \mathrm{Tel}_{\text {children }}-\Delta \mathrm{Tel}_{\text {parent }} \tag{6.10}
\end{equation*}
$$

where:

$$
\begin{aligned}
\mathrm{eTL} \mathrm{XpYp} & =\text { expected telomere length for } \mathrm{XpYp} \\
\mathrm{eTL}_{12 q} & =\text { expected telomere length for } 12 \mathrm{q} \\
\mathrm{X} & =\text { age } \\
\mathrm{oTL}_{\mathrm{XpYp}} & =\text { observed telomere length for } \mathrm{XpYp} \text { measured by STELA } \\
o \mathrm{oTL}_{12 \mathrm{q}} & =\text { observed telomere length for } 12 \mathrm{q} \text { measured by STELA } \\
\Delta \mathrm{Tel}_{\text {children }} & =\text { age-adjusted telomere length for children } \\
\Delta \mathrm{Tel}_{\text {parent }} & =\text { age-adjusted telomere length for parent }
\end{aligned}
$$

Tables 6.4 (for controls) and 6.5 (for LS) show a summary for sex, age, age-adjusted telomere length ( $\Delta \mathrm{TL}$ ) for XpYp and $12 q$ and parent/children pair ( $\mathrm{P} / \mathrm{C}$ ) comparisons that were calculated by subtracting the $\Delta \mathrm{TL}$ for each parent to the $\Delta \mathrm{TL}$ of their offspring as described in equation 6.10.

| Parent |  |  |  |  | Children |  |  |  |  | $\Delta \mathrm{C}-\Delta \mathrm{P}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Code | age | Sex | $\Delta \mathrm{XpYp}$ | -12q | Code | age | Sex | $\Delta \mathrm{XpYp}$ | -12q | XpYp | 12q |
| TEL-F1 G1P1 | 81.6 | F | -1476 | -807 | TEL-F1 G2P1 | 47.1 | F | 528 | -417 | 2004 | 390 |
| TEL-F1 G1P1 | 81.6 | F | -1476 | -807 | TEL-F1 G2P2 | 44.5 | M | -1764 | -221 | -289 | 586 |
| TEL-F1 G1P1 | 81.6 | F | -1476 | -807 | TEL-F1 G2P3 | 43.0 | F | 1028 | -32 | 2504 | 774 |
| TEL-F1 G1P2 | 82.6 | M | -245 | -436 | TEL-F1 G2P1 | 47.1 | F | 528 | -417 | 773 | 19 |
| TEL-F1 G1P2 | 82.6 | M | -245 | -436 | TEL-F1 G2P2 | 44.5 | M | -1764 | -221 | -1519 | 214 |
| TEL-F1 G1P2 | 82.6 | M | -245 | -436 | TEL-F1 G2P3 | 43.0 | F | 1028 | -32 | 1273 | 403 |
| TEL-F2 G1P1 | 54.3 | M | 170 | 514 | TEL-F2 G2P1 | 19.3 | M | -1207 | 285 | -1376 | -229 |
| TEL-F2 G1P2 | 55.2 | F | -2418 | -965 | TEL-F2 G2P1 | 19.3 | M | -1207 | 285 | 1211 | 1250 |
| TEL-F3 G1P1 | 59.8 | M | -216 | 53 | TEL-F3 G2P1 | 22.9 | F | 595 | 990 | 811 | 936 |
| TEL-F3 G1P1 | 59.8 | M | -216 | 53 | TEL-F3 G2P2 | 18.2 | F | 1663 | -745 | 1879 | -799 |
| TEL-F3 G1P2 | 58.4 | F | 3243 | 749 | TEL-F3 G2P1 | 22.9 | F | 595 | 990 | -2648 | 240 |
| TEL-F3 G1P2 | 58.4 | F | 3243 | 749 | TEL-F3 G2P2 | 18.2 | F | 1663 | -745 | -1580 | -1495 |
| TEL-F4 G1P1 | 52.9 | M | 247 | 1151 | TEL-F4 G2P1 | 19.4 | M | -80 | 421 | -327 | -730 |
| TEL-F4 G1P2 | 48.0 | F | 264 | -140 | TEL-F4 G2P1 | 19.4 | M | -80 | 421 | -344 | 561 |
| TEL-F6 G1P2 | 57.5 | F | 119 | -1863 | TEL-F6 G2P1 | 34.4 | F | -1575 | -197 | -1694 | 1666 |
| TEL-F7 G1P1 | 63.8 | M | 612 | 1213 | TEL-F7 G2P1 | 31.4 | M | -380 | -1005 | -991 | -2217 |
| TEL-F7 G1P2 | 62.4 | F | -555 | -785 | TEL-F7 G2P1 | 31.4 | M | -380 | -1005 | 176 | -220 |
| TEL-F8 G1P1 | 53.6 | F | -1305 | -1114 | TEL-F8 G2P1 | 19.9 | F | -1399 | -1613 | -95 | -499 |
| TEL-F8 G1P1 | 53.6 | F | -1305 | -1114 | TEL-F8 G2P2 | 23.6 | F | 368 | 655 | 1673 | 1769 |
| TEL-F8 G1P1 | 53.6 | F | -1305 | -1114 | TEL-F8 G2P3 | 27.7 | M | -169 | -183 | 1136 | 931 |
| TEL-F9 G1P1 | 69.4 | F | 1164 | 2041 | TEL-F9 G2P1 | 37.0 | F | -171 | 440 | -1335 | -1600 |
| TEL-F9 G1P1 | 69.4 | F | 1164 | 2041 | TEL-F9 G2P2 | 45.5 | F | 359 | 697 | -805 | -1344 |
| TEL-F9 G1P1 | 69.4 | F | 1164 | 2041 | TEL-F9 G2P3 | 47.9 | F | -470 | -1026 | -1634 | -3067 |
| TEL-F9 G1P2 | 72.2 | M | 1065 | 990 | TEL-F9 G2P1 | 37.0 | F | -171 | 440 | -1236 | -549 |
| TEL-F9 G1P2 | 72.2 | M | 1065 | 990 | TEL-F9 G2P2 | 45.5 | F | 359 | 697 | -705 | -293 |
| TEL-F9 G1P2 | 72.2 | M | 1065 | 990 | TEL-F9 G2P3 | 47.9 | F | -470 | -1026 | -1534 | -2016 |
| TEL-F11 G1P1 | 62.1 | M | 1120 | -974 | TEL-F11 G2P1 | 29.6 | M | -199 | -678 | -1319 | 296 |
| TEL-F11 G1P1 | 62.1 | M | 1120 | -974 | TEL-F11 G2P2 | 25.6 | F | 1113 | -269 | -6 | 705 |
| TEL-F11 G1P2 | 57.7 | F | 285 | 4 | TEL-F11 G2P1 | 29.6 | M | -199 | -678 | -484 | -682 |
| TEL-F11 G1P2 | 57.7 | F | 285 | 4 | TEL-F11 G2P2 | 25.6 | F | 1113 | -269 | 828 | -273 |
| TEL-F12 G1P1 | 47.2 | F | -150 | 826 | TEL-F12 G2P2 | 19.8 | F | 179 | 1042 | 330 | 216 |
| TEL-F12 G1P1 | 47.2 | F | -150 | 826 | TEL-F12 G2P4 | 23.0 | F | 373 | 839 | 524 | 13 |
| TEL-F12 G1P2 | 53.0 | M | 793 | 217 | TEL-F12 G2P2 | 19.8 | F | 179 | 1042 | -614 | 825 |
| TEL-F12 G1P2 | 53.0 | M | 793 | 217 | TEL-F12 G2P4 | 23.0 | F | 373 | 839 | -420 | 622 |

[^1]Red colour indicates P/C comparisons where children have shorter age-adjusted telomere length than their parents. Table code: Sex : $\mathrm{F}=\mathrm{female}$; $\mathrm{M}=$ male

| Parent |  |  |  |  |  |  |  | Children |  |  |  |  |  |  |  | $\Delta \mathbf{C}-\Delta \mathbf{P}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Code | Age | Gene | Sex | Cancer | Age C. | $\Delta \mathrm{XpYp}$ | -12q | Code | Age | Gene | Sex | Cancer | Age C. | $\Delta \mathrm{XpYp}$ | -12q | XpYp | 12q |
| LRIF1 G2P3 | 68.0 | 2 | M | NO | N/A | -381 | 189 | LRIF1 G3P3 | 43.0 | 2 | F | CRC | 31 | -151 | -863 | 230 | -1051 |
| LRIF1 G2P3 | 68.0 | 2 | M | NO | N/A | -381 | 189 | LRIF1 G3P5 | 41.0 | 2 | F | NO | 33 | -1449 | -180 | -1068 | -368 |
| LRIF1 G2P3 | 68.0 | 2 | M | NO | N/A | -381 | 189 | LRIF1 G3P7 | 38.3 | 2 | F | NO | N/A | -682 | 760 | -301 | 571 |
| LRIF1 G2P3 | 68.0 | 2 | M | NO | N/A | -381 | 189 | LRIF1 G3P8 | 34.7 | 0 | M | NO | N/A | 495 | 927 | 876 | 738 |
| LRIF1 G2P3 | 68.0 | 2 | M | NO | N/A | -381 | 189 | LRIF1 G3P9 | 33.9 | 0 | M | NO | N/A | 323 | -192 | 703 | -380 |
| LRIF1 G2P4 | 66.5 | 0 | F | NO | N/A | 767 | 494 | LRIF1 G3P3 | 43.0 | 2 | F | CRC | 31 | -151 | -863 | -917 | -1357 |
| LRIF1 G2P4 | 66.5 | 0 | F | NO | N/A | 767 | 494 | LRIF1 G3P5 | 41.0 | 2 | F | NO | 33 | -1449 | -180 | -2216 | -674 |
| LRIF1 G2P4 | 66.5 | 0 | F | NO | N/A | 767 | 494 | LRIF1 G3P7 | 38.3 | 2 | F | NO | N/A | -682 | 760 | -1449 | 266 |
| LRIF1 G2P4 | 66.5 | 0 | F | NO | N/A | 767 | 494 | LRIF1 G3P8 | 34.7 | 0 | M | NO | N/A | 495 | 927 | -272 | 432 |
| LRIF1 G2P4 | 66.5 | 0 | F | NO | N/A | 767 | 494 | LRIF1 G3P9 | 33.9 | 0 | M | NO | N/A | 323 | -192 | -444 | -686 |
| LRIF4 G4P8 | 66.7 | 0 | M | NO | N/A | 1388 | 324 | LRIF4 G5P4 | 41.8 | 0 | F | NO | N/A | 164 | -419 | -1224 | -743 |
| LRIF4 G4P9 | 64.1 | 2 | F | NO | N/A | -612 | -149 | LRIF4 G5P4 | 41.8 | 0 | F | NO | N/A | 164 | -419 | 776 | -270 |
| LRIF5 G2P3 | 57.2 | 2 | M | CRC | 47 | -375 | -374 | LRIF5 G3P1 | 27.1 | 2 | F | NO | N/A | -287 | -592 | 89 | -218 |
| LRIF5 G2P3 | 57.2 | 2 | M | CRC | 47 | -375 | -374 | LRIF5 G3P2 | 29.4 | 0 | F | NO | N/A | 99 | -756 | 475 | -382 |
| LRIF7 G4P2 | 58.5 | 2 | M | CRC | 36 | 71 | 850 | LRIF7 G5P2 | 26.4 | 2 | F | MM | 26 | 437 | 557 | 366 | -293 |
| LRIF7 G4P3 | 71.5 | 0 | F | NO | N/A | 895 | 623 | LRIF7 G5P2 | 26.4 | 2 | F | MM | 26 | 437 | 557 | -458 | -66 |
| LRIF12 G3P1 | 61.9 | 0 | M | NO | N/A | -311 | 3 | LRIF12 G4P1 | 31.2 | 1 | F | NO | N/A | -685 | 100 | -373 | 98 |
| LRIF12 G3P1 | 61.9 | 0 | M | NO | N/A | -311 | 3 | LRIF12 G4P2 | 32.6 | 1 | F | NO | N/A | -411 | -485 | -99 | -488 |
| LRIF12 G3P2 | 63.4 | 1 | F | EC | 52 | 116 | 355 | LRIF12 G4P1 | 31.2 | 1 | F | NO | N/A | -685 | 100 | -801 | -255 |
| LRIF12 G3P2 | 63.4 | 1 | F | EC | 52 | 116 | 355 | LRIF12 G4P2 | 32.6 | 1 | F | NO | N/A | -411 | -485 | -527 | -840 |
| LRIF14 G3P2 | 69.9 | 0 | M | NO | N/A | -966 | -532 | LRIF14 G4P2 | 45.7 | 2 | F | NO | N/A | -812 | -360 | 154 | 173 |
| LRIF14 G3P3 | 66.9 | 2 | F | NO | N/A | 336 | -162 | LRIF14 G4P2 | 45.7 | 2 | F | NO | N/A | -812 | -360 | -1148 | -197 |
| BWHF1a G3P1 | 64.4 | 1 | M | SCC | 45 | -286 | 535 | BWHF1a G4P3 | 34.6 | 1 | F | CRC | 14 | -127 | 695 | 160 | 160 |
| BWHF1a G3P2 | 63.6 | 0 | F | NO | N/A | -128 | 153 | BWHF1a G4P3 | 34.6 | 1 | F | CRC | 14 | -127 | 695 | 1 | 541 |
| BWHF3a G3P2 | 46.9 | 1 | F | CRC | 33 | 563 | -65 | BWHF3a G4P1 | 28.8 | 1 | M | NO | N/A | 182 | 1656 | -381 | 1721 |
| BWHF3a G3P2 | 46.9 | 1 | F | CRC | 33 | 563 | -65 | BWHF3a G4P3 | 27.6 | 0 | F | NO | N/A | 2727 | 480 | 2164 | 544 |
| BWHF5a G3P1 | 57.6 | 0 | F | BCC | 52 | 388 | -640 | BWHF5a G4P6 | 23.3 | 0 | F | NO | N/A | -369 | -1003 | -757 | -363 |


| Parent |  |  |  |  |  |  |  | Children |  |  |  |  |  |  |  | $\Delta \mathbf{C}-\Delta \mathbf{P}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Code | Age | Gene | Sex | Cancer | Age C. | $\Delta \mathrm{XpYp}$ | $\Delta 12 \mathrm{q}$ | Code | Age | Gene | Sex | Cancer | Age C. | $\Delta \mathrm{XpYp}$ | $\Delta 12 \mathrm{q}$ | XpYp | 12q |
| BWHF6a G3P1 | 63.3 | 1 | F | EC | 52 | 1231 | -402 | BWHF6a G4P2 | 31.9 | 1 | F | NO | N/A | -56 | -99 | -1286 | 303 |
| BWHF6a G3P1 | 63.3 | 1 | F | EC | 52 | 1231 | -402 | BWHF6a G4P3 | 36.4 | 1 | M | NO | N/A | -699 | -1550 | -1930 | -1148 |
| BWHF11a G3P1 | 62.4 | 2 | M | CRC | 60 | 795 | -488 | BWHF11a G4P4 | 38.2 | 2 | M | NO | N/A | -463 | -173 | -1258 | 315 |
| BWHF2 G3P7 | 66.9 | 1 | M | CRC | 38 | 1504 | 1443 | BWHF2 G4P11 | 39.3 | 1 | M | CRC | 23 | -1054 | -417 | -2558 | -1860 |
| BWHF2 G3P8 | 66.4 | 0 | F | NO | N/A | -1443 | -706 | BWHF2 G4P11 | 39.3 | 1 | M | CRC | 23 | -1054 | -417 | 389 | 289 |
| BWHF42 G5P2 | 64.8 | 1 | F | CRC | 51 | 363 | 1009 | BWHF42 G6P1 | 34.9 | 1 | M | CRC | 20 | 246 | 166 | -117 | -843 |
| BWHF42 G5P2 | 64.8 | 1 | F | CRC | 51 | 363 | 1009 | BWHF42 G6P2 | 32.8 | 0 | M | NO | N/A | 1697 | 348 | 1333 | -661 |
| BWHF42 G5P2 | 64.8 | 1 | F | CRC | 51 | 363 | 1009 | BWHF42 G6P3 | 28.3 | 0 | F | NO | N/A | -1279 | 132 | -1643 | -878 |
| BWHF55 G3P1 | 58.2 | 0 | M | NO | N/A | -1044 | -1048 | BWHF55 G4P1 | 35.1 | 2 | M | NO | N/A | -856 | -614 | 188 | 435 |
| BWHF55 G3P1 | 58.2 | 0 | M | NO | N/A | -1044 | -1048 | BWHF55 G4P2 | 31.8 | 2 | F | NO | N/A | 116 | -309 | 1161 | 739 |
| BWHF55 G3P2 | 58.8 | 2 | F | CRC | 31 | -358 | 86 | BWHF55 G4P1 | 35.1 | 2 | M | NO | N/A | -856 | -614 | -498 | -700 |
| BWHF55 G3P2 | 58.8 | 2 | F | CRC | 31 | -358 | 86 | BWHF55 G4P2 | 31.8 | 2 | F | NO | N/A | 116 | -309 | 474 | -395 |
| BWHF55 G3P3 | 55.5 | 2 | M | CRC | 31 | 272 | 337 | BWHF55 G4P4 | 19.2 | 2 | M | CRC | 15 | -3029 | -2296 | -3301 | -2633 |
| BWHF71 G3P1 | 66.2 | 0 | M | NO | N/A | -504 | -147 | BWHF71 G4P1 | 36.1 | 0 | M | NO | N/A | 413 | -32 | 916 | 116 |
| BWHF71 G3P1 | 66.2 | 0 | M | NO | N/A | -504 | -147 | BWHF71 G4P2 | 34.4 | 2 | F | NO | N/A | -157 | 579 | 347 | 726 |
| BWHF71 G3P2 | 65.9 | 2 | F | EC | 50 | -788 | -75 | BWHF71 G4P1 | 36.1 | 0 | M | NO | N/A | 413 | -32 | 1201 | 44 |
| BWHF71 G3P2 | 65.9 | 2 | F | EC | 50 | -788 | -75 | BWHF71 G4P2 | 34.4 | 2 | F | NO | N/A | -157 | 579 | 631 | 654 |
| BWHF100 G4P1 | 53.6 | 2 | M | CRC | 44 | -679 | 700 | BWHF100 G5P1 | 26.9 | 2 | F | NO | N/A | -1925 | 361 | -1246 | -339 |
| BWHF100 G4P2 | 52.6 | 0 | F | NO | N/A | -606 | -467 | BWHF100 G5P1 | 26.9 | 2 | F | NO | N/A | -1925 | 361 | -1319 | 828 |
| BWHF102 G3P3 | 61.8 | 1 | M | CRC | 40 | 882 | 1260 | BWHF102 G4P4 | 30.9 | 1 | F | NO | N/A | 512 | -1359 | -370 | -2619 |
| BWHF105 G3P5 | 64.1 | 2 | F | EC | 31 | 459 | 127 | BWHF105 G4P2 | 41.0 | 2 | F | UBC | 33 | -326 | -837 | -786 | -963 |
| BWHF105 G3P5 | 64.1 | 2 | F | EC | 31 | 459 | 127 | BWHF105 G4P4 | 38.8 | 0 | F | NO | N/A | 1083 | 352 | 624 | 225 |
| BWHF105 G3P6 | 67.6 | 2 | F | CRC | 60 | 412 | -751 | BWHF105 G4P9 | 42.0 | 2 | F | NO | N/A | 9 | -319 | -403 | 432 |
| BWHF105 G3P6 | 67.6 | 2 | F | CRC | 60 | 412 | -751 | BWHF105 G4P11 | 37.1 | 0 | M | NO | N/A | 4 | 856 | -408 | 1607 |
| BWHF105 G4P9 | 42.0 | 2 | F | NO | N/A | 9 | -319 | BWHF105 G5P4 | 20.6 | 2 | F | NO | N/A | 447 | -930 | 438 | -611 |
| BWHF106 G4P2 | 51.0 | 0 | F | NO | N/A | -1644 | 532 | BWHF106 G5P1 | 27.8 | 1 | M | NO | N/A | -599 | -914 | 1046 | -1446 |
| BWHF106 G4P1 | 58.2 | 1 | M | CRC | 44 | -873 | 845 | BWHF106 G5P1 | 27.8 | 1 | M | NO | N/A | -599 | -914 | 275 | -1759 |


| Parent |  |  |  |  |  |  |  | Children |  |  |  |  |  |  |  | $\Delta \mathbf{C}-\Delta \mathbf{P}$ |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Code | Age | Gene | Sex | Cancer | Age C. | $\Delta \mathrm{XpYp}$ | -12q | Code | Age | Gene | Sex | Cancer | Age C. | $\Delta \mathrm{XpYp}$ | -12q | XpYp | 12 q |
| BWHF107 G3P3 | 72.4 | 1 | M | CRC | 57 | -585 | -475 | BWHF107 G4P1 | 48.0 | 1 | M | NO | N/A | -1449 | -473 | -864 | 2 |
| BWHF107 G3P3 | 72.4 | 1 | M | CRC | 57 | -585 | -475 | BWHF107 G4P2 | 46.3 | 0 | F | NO | N/A | -592 | -956 | -7 | -481 |
| BWHF107 G3P4 | 73.0 | 0 | F | TC | 65 | 357 | 772 | BWHF107 G4P1 | 48.0 | 1 | M | NO | N/A | -1449 | -473 | -1806 | -1244 |
| BWHF107 G3P4 | 73.0 | 0 | F | TC | 65 | 357 | 772 | BWHF107 G4P2 | 46.3 | 0 | F | NO | N/A | -592 | -956 | -949 | -1728 |
| BWHF108 G3P3 | 73.1 | 2 | M | UBC | 64 | -512 | -302 | BWHF108 G4P7 | 37.7 | 2 | F | NO | N/A | 871 | 1039 | 1383 | 1341 |
| BWHF108 G3P3 | 73.1 | 2 | M | UBC | 64 | -512 | -302 | BWHF108 G4P8 | 30.7 | 2 | M | NO | N/A | 261 | 369 | 773 | 671 |
| BWHF108 G3P4 | 71.9 | 0 | F | NO | N/A | -7 | 660 | BWHF108 G4P7 | 37.7 | 2 | F | NO | N/A | 871 | 1039 | 879 | 379 |
| BWHF108 G3P4 | 71.9 | 0 | F | NO | N/A | -7 | 660 | BWHF108 G4P8 | 30.7 | 2 | M | NO | N/A | 261 | 369 | 268 | -290 |
| BWHF125 G4P10 | 62.1 | 0 | M | NO | N/A | -115 | -607 | BWHF125 G5P2 | 25.7 | 2 | M | NO | N/A | -258 | -1041 | -142 | -434 |
| BWHF125 G4P10 | 62.1 | 0 | M | NO | N/A | -115 | -607 | BWHF125 G5P3 | 22.9 | 2 | F | NO | N/A | -2015 | -2219 | -1899 | -1612 |
| BWHF125 G4P9 | 49.0 | 2 | F | CRC | 34 | -867 | -526 | BWHF125 G5P2 | 25.7 | 2 | M | NO | N/A | -258 | -1041 | 609 | -516 |
| BWHF125 G4P9 | 49.0 | 2 | F | CRC | 34 | -867 | -526 | BWHF125 G5P3 | 22.9 | 2 | F | NO | N/A | -2015 | -2219 | -1148 | -1694 |
| BWHF126 G3P6 | 65.3 | 1 | F | EC | 64 | 555 | -20 | BWHF126 G4P3 | 40.1 | 1 | M | CRC | 29 | 64 | 979 | -492 | 999 |
| BWHF126 G3P6 | 65.3 | 1 | F | EC | 64 | 555 | -20 | BWHF126 G4P5 | 42.7 | 0 | F | NO | N/A | 568 | 72 | 13 | 92 |
| BWHF126 G3P6 | 65.3 | 1 | F | EC | 64 | 555 | -20 | BWHF126 G4P6 | 32.3 | 1 | M | NO | N/A | 1059 | -62 | 503 | -42 |
| BWHF130 G4P1 | 53.2 | 1 | M | CRC | 46 | 1923 | -234 | BWHF130 G5P3 | 23.4 | 1 | M | NO | N/A | -1262 | -475 | -3185 | -240 |
| BWHF130 G4P2 | 49.9 | 0 | F | NO | N/A | 2362 | 1545 | BWHF130 G5P3 | 23.4 | 1 | M | NO | N/A | -1262 | -475 | -3624 | -2020 |
| BWHF130 G4P5 | 49.3 | 1 | M | CRC | 26 | 785 | -308 | BWHF130 G5P9 | 23.7 | 0 | M | NO | N/A | -747 | 1662 | -1532 | 1970 |
| BWHF130 G4P4 | 50.4 | 0 | M | NO | N/A | -732 | 673 |  |  |  |  |  |  |  |  |  |  |
| BWHF130 G4P7 | 44.8 | 1 | M | CRC | 33 | 506 | -491 |  |  |  |  |  |  |  |  |  |  |

[^2]Figure 6.12.A shows the age-adjusted telomere length ( $\Delta \mathrm{Tel}$ ) for all controls ( 52 samples) and LS ( 76 samples). No differences were observed in the distribution of the age-adjusted telomere lengths for either XpYp or $12 q$ between the control and the LS cohort individuals. However, when dividing by the genotype, it was found that $\mathrm{MSH}^{+/-}$individuals had significantly shorter $\Delta \mathrm{XpYp}$ than the $\mathrm{MMR}^{+/+}$children in the LS cohort, suggesting that $\mathrm{MSH}^{+/-}$patients might have shorter XpYp but not 12 q telomere than MMR proficient patients, although not very different from controls. Additionally, when comparing the rank between parents and children for both cohorts using two-tailed Mann-Whitney test we found no differences in controls for XpYp ( $p=0.6$ ) or $12 q(p=0.8)$ but significantly shorter $\operatorname{XpYp}(p=0.0169)$ and $12 q$ ( $\mathrm{p}=0.0485$ ) telomeres in the LS children compared to the LS parents (figure 6.12.B. Subsequently, parent/children pair comparisons were performed to calculate the change in telomere length as explained in equation 6.10 and no differences were observed between the two cohorts. However, and despite the very small sample size ( 7 pairs), in the MMR deficient pairs that have developed cancer in both generations, children seemed to have shorter telomeres than their parents (figure 6.12.C).

Interestingly, when the P/C pair comparison within the LS were studied accounting for the genotype in both generations, a similar trend for both telomeres (figure 6.13) was found, suggesting that a) $\mathrm{MMR}^{+/-}$children had shorter mean age-adjusted telomere length ( $\Delta \mathrm{Tel}$ ) than their parents (independent of the genotype); b) the $\mathrm{MMR}^{+/+}$children in the LS cohort had longer mean $\Delta \mathrm{Tel}$ than their $\mathrm{MMR}^{+/-}$parents and siblings and c) $\mathrm{MMR}^{+/+}$children in the LS cohort had shorter mean $\Delta \mathrm{Tel}$ than their $\mathrm{MMR}^{+/+}$parent. Due to this final unexpected finding, it was concluded that despite not inheriting the germline mutation, having a $\mathrm{MMR}^{+/-}$parent might have an effect on the WT offspring's telomeres.

Telomere length was very heterogeneous among individuals with the same age, specially in the LS cohort as previously shown by lower r values for the regression analysis compared to the control cohort. Thus, parent/children pair comparisons might be very different due to that variability even between


Figure 6.12: Age-adjusted telomere length. A shows the mean age-adjusted telomere length ( $\Delta$ Tel) for controls (red) and LS (blue for XpYp and green for $12 q$ telomeres). B compares the age-adjusted telomere lengths ( $\Delta \mathrm{Tel}$ ) between parent and children within the two cohorts. The XpYp telomere is shown in blue colours and the $12 q$ telomere is shown in green colours (for both, darker and lighter colours are used for parents and children respectively). C shows the parent/children pair comparisons for $\Delta T e l$ in controls (red) and LS (purple). In addition, 7 pairs are shown separately for $\mathrm{MMR}^{+/-}$parents and children that have both developed cancer (yellow). " n " indicates the number of pairs analysed per group. Means are shown as horizontal black lines and $p$ values for two tailed-Mann-Whitney test are shown for A and B .


Figure 6.13: Change in telomere length over generations. The scatter plots show child/parent comparisons for $\mathrm{XpYp}(\mathbf{A})$ and $12 q(\mathbf{B})$ telomeres. Controls are shown in red colour; $\mathrm{MMR}^{+/-}$in both generations are shown in purple; $\mathrm{MMR}^{+/+}$in both generations in blue; $\mathrm{MMR}^{+/-}$children compared to their $\mathrm{MMR}^{+/+}$parent in yellow; $\mathrm{MMR}^{+/+}$children compared to their $\mathrm{MMR}^{+/-}$parent in dark green and $\mathrm{MMR}^{+++}$children compared to their $\mathrm{MMR}^{+/-}$siblings in light green. Means are shown as horizontal black lines. " $n$ " indicates the number of pairs analysed per group.
pairs belonging to the same group (see large distribution in figure 6.13). Consequently, we decided to study the P/C pair comparisons within each family to find a difference in the pattern of telomere length inheritance. In addition, other factors affecting TL like environmental were also removed (as it would be expected that members from the same family would live under similar environmental conditions, at least during childhood when the telomere shortening rates have been reported to be greater Frenck et al. (1998) and Rufer et al. (1999)). Data from tables 6.4 and 6.5 were used to plot graphs in figure 6.14. Considering that the number of comparison per family were a limitation, finding a trend grouping the change in TL over the generations for the control families was difficult; an example of that great heterogeneity was shown by two families exhibiting the same number of comparisons (TELF1 and TELF9) that nevertheless behaved very differently. On the one hand, in TELF9 all three daughters had shorter XpYp and 12q telomeres compared to both parents while in TELF1, both daughters had longer telomeres than their parents and only the son (TELF1G2P2) had shorter XpYp than both parents. Interestingly, there were three sons (TELF2G2P1, TELF4G2P1 and TELF7G2P1) with shorter telomeres than their respective fathers but not their mothers. Conversely, son TELF11G2P1 had both telomeres shorter than his mother and only shorter XpYp than his father. In family TELF3, both daughters had shorter XpYp telomere than their mother but longer than their father; on the contrary, for family TELF12, both daughters had shorter XpYp telomere than their father but not their mother (it is noteworthy that in those two families, the mother -for TELF3- and the father -for TELF12were cancer-affected -breast and testicular cancer respectively-). Additionally, daughter TELF3G2P2 had shorter 12q than both parents whereas her sister had longer $12 q$ than both parents. For family TELF8, only the youngest child had shorter telomeres than her mother compared to her two siblings. Finally, both children in family TELF11 had longer 12q telomere than their father but not their mother.

For the LS cohort, parent/children pair comparisons per family are shown in figure 6.14.B. Contrary to the control families (figure 6.14.A), there seemed
to be a tendency to smaller variations between parent/children pairs, as there were several families (LRIF5, LRIF7, LRIF14, BWHF55-A, BWHF100, BWHF 105A, BWHF105C, BWHF5a; for MSH2 and LRIF12, BWHF126 and BWHF1a; for MLH1) having values for $\Delta$ Children- $\Delta$ Parent very close to zero, suggesting that in those families, telomeres did not change in length between the generations. Conversely, there were some $\mathrm{P} / \mathrm{C}$ comparisons that showed very negative $\Delta$ Children- $\Delta$ Parent values (children with smaller age-adjusted telomeres than their parents) and only three comparisons (in BWF130-B and BWHF3a MLH1 families) that showed greater values (children with longer telomeres than their parents). Interestingly, individual BWHF55G4P4 had the shortest age-adjusted telomeres (both XpYp and $12 q$ ) and an earlier age onset of CRC ( 15 years) compared to his father ( 31 years). Also in family BWHF2, the individual G4P11 that had developed CRC at 23 years had much shorter telomeres than his father (who developed CRC at 38 years). Nevertheless, telomeres in this individual were slightly longer than his $\mathrm{MMR}^{+/+}$ mother. Other P/C pair comparisons between cancer-affected pairs (in families LRIF7, BWHF42, BWHF126 and BWHF1a) did not show very negative $\Delta$ Children- $\Delta$ Parent values, suggesting that cancer in the second generation might not be driven by very short telomeres.

B



| $\Delta 12 \mathrm{MMR}^{+/-}$child $-\Delta 12 \mathrm{qMMR}^{+/-}$parent | - $\Delta 12 \mathrm{qMMR}^{+/-}$child $-\Delta 12 \mathrm{qMMR}{ }^{+/+}$parent |
| :---: | :---: |
| $\Delta \mathrm{XpYpMMR}^{+/-}$child $-\Delta \mathrm{Xp}_{\text {PlpMMR }}{ }^{+/-}$parent |  |
| - $\Delta 12 \mathrm{qMMR}{ }^{+/+}$child $-\Delta 12 \mathrm{qMMR}^{+/-}$parent | - $\Delta 12 \mathrm{qMMR}{ }^{+/+}$child $-\Delta 12 \mathrm{qMMR}{ }^{+/+}$parent |
| $\triangle$ Хp Yp |  |

Figure 6.14: Comparison of median telomere lengths within families. A shows a bar graph for control families. Blue colour indicates XpYp telomere and green, 12q. All parent-children comparisons are shown indicating the pairs in the X axis. $\mathbf{B}$ shows two bar graphs for the LS families. The top graph shows only MSH2 families and the bottom graph shows only MLH1 families. Red asterisks indicate MMR-MMR pairs with both cancer-affected generations. The colour code is indicated below the second graph.

### 6.4 Discussion

Saliva samples have been reported to contain approximately $4.3 \times 10^{5}$ cells $/ \mathrm{ml}$, consisting of either epithelial cells or leukocytes (Chiappin et al., 2007) although in different concentrations depending on the health status of the donor (Thiede et al., 2000). In addition, saliva contains hormones and a large amount of proteins that varies accounting for the glandular activity (reviewed in Chiappin et al. (2007)), and also other contaminants such as bacterial DNA ( $\approx$ $1.7 \times 10^{7}$ bacteria $/ \mathrm{ml}$ ) or food debris that might contribute to the variability between samples (Sun \& Reichenberger, 2014). Nevertheless, this type of biological sample is increasing its popularity as diagnostic tool as it is a non-invasive, cheap way of obtaining reasonable amounts of DNA compared to blood samples (reviewed in Sun \& Reichenberger (2014)). Telomere length has been measured in saliva samples using Q-PCR in several cross-sectional studies ((Hewakapuge et al., 2008), (Theall et al., 2013), and (Chen et al., 2015)) and it has been reported that telomeres tend to be longer in saliva samples than in blood as the cell turnover is lower in the former ((Theall et al., 2013) and (Mitchell et al., 2014)).

Since its development in 2003, STELA (single telomere length analysis) has been used to measure telomere length in a variety of human primary cell lines, mainly fibroblasts ((Baird et al., 2003), (Britt-Compton et al., 2006) and (Mendez-Bermudez \& Royle, 2011)) and blood cells ((Hills et al., 2009) and (Britt-Compton et al., 2009a)) and a strong correlation between telomere length measured by STELA and TRF (Terminal Restriction Fragments) has been described ((Baird et al., 2003) and (Baird et al., 2006)). Nevertheless, it has not being used, at least to our knowledge, to measure telomere length in saliva samples. In Lin et al. (2010), STELA was used to measure telomere length in 41 blood samples from chronic lymphocytic leukaemia (CLL) patients (that are characterised by very short telomeres) with an age-range of 50 to 89 years (median = 71 years). Later in Lin et al. (2014), STELA was used in two bigger cohorts of CLL blood samples: one with an age range of 27 to 95 years (median $=64$ years) and another with an age-range of 35 to 90
years (median $=66$ years). Taking into account that the age-range in our cohorts was 18 to 83 years (for controls) and 19 to 73 years (for Lynch syndrome) and that we used a tissue for DNA extraction with a reported longer telomere length compared to blood, an underestimation of the TL in the younger individuals was not unexpected.

Telomere length for XpYp and $12 q$ telomeres was measured by STELA in 128 saliva samples from control (52) and Lynch syndrome (76) individuals to study whether inheritance of a germline mutation in a DNA mismatch repair (MMR) gene (either $h M S H 2$ or $h M L H 1$ ) had any impact on telomere length dynamics. A significant relationship between telomere length and age was only found when data for individuals younger than 35 years was transformed, possibly due to an underestimation of TL as explained before. After data transformation, XpYp and 12 q telomeres shortened at similar, although slightly higher, rates in controls ( 34 and $31 \mathrm{bp} /$ year for XpYp and 12 q respectively) than in LS patients ( 22 and $19 \mathrm{bp} /$ year for XpYp and 12q respectively).

We found that some children in control families had shorter telomeres than their parents ( $42 \%$ and $26 \%$ for XpYp and $12 q$ respectively), although the percentage decreased after data transformation. On the one hand, this might be explained if telomeres were shortening at a faster rate between 18 to 35 years than after 35 years, what might be supported by Frenck et al. (1998) who demonstrated that after a rapid decline in telomere length from birth until the age of 4 years, there was a stabilisation in the telomere length between the ages of 10 and 30 in human leukocytes that was followed by a more modest telomere attrition rate with age. Being that the case in our cohort, the constant used to transform the data should had been greater. On the other hand, the assumption that children have longer telomeres than their parents comes from cross-sectional studies where the telomere length of all individuals from different ages fits to a linear regression according to the telomere attrition rate estimated. Nevertheless, some longitudinal studies available at the moment (in which telomere length is measured for the same individual at least in two time points separated by a certain time period, either months or years), suggest that there is not only telomere loss at different rates between
individuals but also either no change or even gain of telomeric repeats during the follow-up period ((Aviv et al., 2009), (Chen et al., 2011), (Steenstrup et al., 2013)). In addition, the strong paternal age effect on the offspring's TL, suggesting that children from older fathers have longer telomeres possibly due to the elongation of the telomere length in sperm with age ((De Meyer et al., 2007), (Kimura et al., 2008) and (Broer et al., 2013)), might contribute to changes in the TL over generations as suggested by Holohan et al. (2015).

Another surprising finding was the fact that the percentage of normality (individuals with either XpYp or 12 q normally distributed telomere length) was lower in LS than controls, which was explained by the presence of many more individuals showing outliers (single telomere lengths measured at low frequency) in the LS cohort. Besides, those outliers were, without exception, long molecules, suggesting that shorter telomeres were more abundant in LS than longer ones. As the telomere attrition rate with age was not different between both cohorts (confirming results in Seguí et al. (2013) using Q-PCR and a bigger cohort), the abundance of short telomeres in LS individuals might be explained by: a) inheritance of shorter telomeres compared to controls or b) faster shortening rate earlier in life or even during embryogenesis. In addition, contrary to what was described by Seguí et al. (2013), our results did not suggest shorter telomeres and faster telomere shortening rate with age in cancer-affected $\mathrm{MMR}^{+/-}$compared to controls and $\mathrm{MMR}^{+/-}$cancer-free patients. In addition, telomeres in MSH2 families were shorter than in MLH1 families in which no significant relationship between TL and age was found, confirming previous results in Bozzao et al. (2011a). Finally, parent/children (P/C) pair comparisons for age-adjusted telomere lengths ( $\Delta$ Tel) showed a similar trend for both telomeres suggesting that $\mathrm{MMR}^{+/-}$children had shorter telomeres than both their $\mathrm{MMR}^{+/-}$and $\mathrm{MMR}^{+/+}$parent. Besides, $\mathrm{MMR}^{+/+}$ children seemed to have longer telomeres than their $\mathrm{MMR}^{+/-}$parent and siblings but not their $\mathrm{MMR}^{+/+}$parent, suggesting that a shorter telomere length might have been inherited from the $\mathrm{MMR}^{+/-}$parent in agreement with Aubert et al. (2012a) who showed that $h T E R T$ or $h T E R C$ gene mutation non-carrier relatives had shorter than expected telomeres in blood.

## 7 Final remarks

The effect of DNA mismatch repair (MMR) deficiency on telomere dynamics has been explored in this Thesis. Particularly, I focused on the loss of the MSH2 protein by shRNA downregulation in a human primary cell line (WI38), in a colorectal cancer cell line (LoVo) with mutated MSH2 gene and in germline mutations for either MSH2 or MLH1 genes in Lynch syndrome patients.

We found that MSH2 downregulation in WI38 cells grown under $5 \% \mathrm{O}_{2}$ increased the heterogeneity in telomere shortening rates and that telomeres, in clones with protein concentrations levels below $40 \%$ had a significantly faster shortening rate than in control clones. Interestingly, the telomere associated with the $12 q B$ haplotype in the telomere adjacent DNA shortened exponentially faster with decreasing protein levels but this linear relationship was not significant for the other 12 q allele ( $\Delta$ ). The interspersion pattern of the telomere variant repeat (TVR) repeat was compared between different telomeres and we found that the $12 q$ had more variant repeats compared to XpYp or 17 p telomeres; additionally, the interspersed repeat pattern between the $12 q$ alleles showed private repeat combinations that might be responsible for the different behaviour of the alleles although experiments on the mutation frequency on those repeats are needed to confirm it. Comparisons of the findings in this Thesis with preliminary experiments suggested that the effect of MMR deficiency might be reinforced by high oxygen concentrations, as cells with depleted MSH2 grown under $21 \% \mathrm{O}_{2}$ had a greater difference in telomere shortening rates compared to controls (Mendez-Bermudez \& Royle, 2011) than in the present experiment performed under biological, $5 \%, \mathrm{O}_{2}$ conditions.

It was previously hypothesised that the telomere instability found in MMRdeficient cancer cells might also contribute to the increase in telomere shortening observed before. Hence, mutant telomeres in LoVo were analysed and a 5:2 ratio for deletion:insertion types was found; however, more molecules
(15) were analysed containing any of the two insertion types compared to the molecules carrying a deletion (10) due to somatic mosaicism. Comparison with previous mutation analysis on the same cell line showed one shared mutation and the combined mutation frequency (5.8\%) indicated a 9:3 ratio for deletions:insertions, suggesting that the former were three times more common than the latter. Therefore, the accumulation of deletions might contribute to the increase in telomere shortening rates described before but also, the loss of some variant repeats involved in the recruitment of the shelterin component TRF2, might be involved in telomere deprotection, although it needs to be confirmed in the future. Additionally, the TVR protocol was optimised to make it suitable for capillary electrophoresis increasing the resolution and throughput of the method.

Furthermore, a new third next generation sequencing technology (minION ${ }^{\text {TM }}$ ) for long reads was tested to explore its applicability for telomere sequencing. Two control amplicons (previously characterised by Enjie Zhang using Ion Torrent ${ }^{\mathrm{TM}}$ ) carrying interspersed homopolymeric repeats were sequenced with a greater than $95 \%$ identity. The subtelomeric region of the XpYp was also sequenced with a more than $93 \%$ identity, with the possibility of distinguishing between alleles in the highest quality reads. Moreover, a telomeric reference based on the TVR map was used to align the reads against it. The variant repeats pattern was not accurately sequenced but a lack of identity to the reference was found in the most distant position (from 4550 bp to the end) in which some variant repeats such as TACGGG, CTAGG or TCAGGG were sequenced, suggesting that a not previously described, more distant, variant repeat region might be present in this telomere. However, further analyses are needed to confirm this and the use of the latest chemistry and device for the nanopore sequencing might improve these results.

Besides, the effect of telomerase inhibition by a small non-nucleosidic drug called BIBR1532 in the LoVo cell line was tested as it has been hypothesised that tumours with very short telomeres might benefit more from this therapy. The drug reduced significantly the cell growth in the control (MMR-proficient) colon cancer cell line but not in LoVo. The cell cycle was not significantly
affected either in LoVo, although it showed an increase in S-phase for the other cell line. Correspondingly, the telomere shortening rate increased significantly in the MMR-proficient cell line treated with the drug but not in LoVo. However, those results were not conclusive as no telomerase activity reduction was observed after the treatment measured by TRAP assay. Ultimately, we hypothesised that, as the elongation function of telomerase is targeted by this chemical, telomerase activity might be intact, therefore regulating transcription and metabolic reprogramming that, despite the effect on telomere shortening rate, might still drive tumorigenesis. Therefore, the use of the DN-hTERT plasmid, that completely degrades the endogenous telomerase, or the site-directed mutagenesis of the hTERC or hTERT genes using CRISPR-Cas9 approaches might give a better insight into the effect of telomerase inhibition on MMR-deficient cells.

Finally, telomere length in saliva samples from control and Lynch syndrome families was studied by STELA and it was found that: a) telomeres shortened with age at a similar rate in both cohorts with no effect based on sex, b) there was no relationship between telomere length and age in $M L H 1^{+/-}$ patients, c) the age-adjusted telomere length for XpYp but not $12 q$ was significantly shorter in $\mathrm{MSH}^{+/-}$carriers compared to $\mathrm{MMR}^{+/+}$children within the LS cohort, d) the $\mathrm{MMR}^{+/-}$children had shorter age-adjusted telomere lengths than both parents and e) the $\mathrm{MMR}^{+/+}$children had longer age-adjusted telomeres than their $\mathrm{MMR}^{+/-}$parents and siblings but shorter than their $\mathrm{MMR}^{+/+}$ parent. The lack of difference in the telomere shortening rate with age between the two cohorts might be due different initial telomere lengths over generations what has been proposed an explanation for the distinct attrition rates found between cross-sectional and longitudinal studies; hence, it might be interesting to consider following up some of the most interesting families over time, in order to understand the direct effect of MMR-deficiency on the telomeres of specific individuals.

## A Materials summary

List of the main materials used in this Thesis. The catalog number and the supplier is also indicated for an easier identification.

|  | Product | Supplier, Country | Catalog number |
| :---: | :---: | :---: | :---: |
| Antibodies | Anti-MSH2 | Abcam, UK | ab52266 |
|  | Anti-GAPDH | Life Technologies, UK | AM4300 |
|  | ECL Mouse IgG from sheep | GE Healthcare, UK | NA931 |
| Cell staining | $\beta$-galactosidase staining kit | Cell Signalling, UK | 9860 |
|  | Propidium Iodide | Sigma-Aldrich, UK | P4170 |
|  | Trypan blue solution 0.4\% | Life-Technologies, UK | 15250-061 |
|  | 4,6-Diamidino-2-phenylindole dihydrochloride, DAPI | Sigma-Aldrich, UK | 32670 |
| Chemicals and others | Formamide | Merck, UK | S4117 |
|  | AccuGel (40\%) 19:1 (acrylamide:Bis acrylamide) | Fisher Scientific, UK | 12301469 |
|  | ProtoGel (30\%) 37.5:1 (acrylamide:Bis acrylamide) | Fisher Scientific, UK | 12381469 |
|  | Cloning discs | Sigma-Aldrich, UK | Z374466100EA |
|  | BLUeye Prestained Protein Ladder | GeneFlow | S6-0024 |
|  | Hi-Di ${ }^{\text {TM }}$ formamide | Applied Biosystems | 4311320 |
|  | MapMarker ${ }^{\circledR}$ X-Rhodamine Labeled 1000 bp | BioVentures, US | MM-1000XL-ROX |
|  | MAGNA nylon transfer membrane 0.45 microns | Fisher Scientific, US | 1213403 |
|  | Minimum Essential Medium (MEM) | Thermo Scientific, UK | 11095-080 |
|  | Dulbecco's Modified Eagle Medium (D-MEM ) | Thermo Scientific, UK | 41965-062 |
|  | 100X Non-Essential AminoAcids (NEAA) | Thermo Scientific, UK | 11140-050 |
|  | Fetal Bovine Serum (FBS) South African | Pan-Biotech, Germany | P30-1502 |
|  | Protran Premium Nitrocellulose membrane | Fisher Scientific, UK | 15269794 |
| Drugs | Puromycin dihydrocloride | Sigma-Aldrich, UK | P9620 |
|  | Ampicillin | Sigma-Aldrich, UK | A0166 |
|  | BIBR1532 | Stratech Scientific, UK | S1186-SEL |
| Enzymes | End-repair using NEBNext ${ }^{\circledR}$ End Repair Module | NEB, UK | E6050S |
|  | A-tailing using NEBNext ${ }^{\circledR}$ dA-Tailing Module | NEB, UK | E6053S |
|  | Blunt/TA Ligase Master Mix | NEB, UK | M0367S |
| Kits | E.Z.N.A. ${ }^{\circledR}$ Plasmid Maxi Kit | Omega | D6922-04 |
|  | MaXtract High Density | Qiagen, UK | 129056 |
|  | Zymoclean ${ }^{\text {™ }}$ Gel DNA Recovery Kit | ZymoResearch, UK | D4002 |
|  | Oragene ${ }^{\text {m/ }}$ DNA saliva kit | Genotek, US | OG-500 |
|  | ECL prime | GE healthcare | RPN2236 |
|  | TRAPeze ${ }^{\text {® }}$ | Millipore, UK | S7700 |
|  | Ingenio ${ }^{\left({ }^{\text {® }}\right.}$ Electroporation Kit | Mirus, UK | MIR 50115 |
|  | Quick Start ${ }^{\text {mm }}$ Bradford Protein Assay Kit | BioRad, UK | 500-0201 |
|  | Pellet Paint ${ }^{\text {® }}$ | Millipore, UK | 69049 |
|  | Genomic DNA kit | Oxford Nanopore, UK | SQK-MAP004 |
|  | AmPure beads | Beckman Coulter, UK | A63880 |
|  | Dynabeads ${ }^{\circledR}$ His-Tag | Thermo Scientific, UK | 10103D |

## B Primers summary

List of the primers used in this Thesis. "T" refers to annealing temperature and "V" indicates variable depending on the forward optimum annealing temperature. Letters in bold indicate the allele-specific nucleotide.

|  | Name | Sequence 5' to 3' | T |
| :---: | :---: | :---: | :---: |
| $\begin{aligned} & \text { 殅 } \\ & \text { E } \end{aligned}$ | Telorette2 | TGCTCCGTGCATCTGGCATCTAACCCT | V |
|  | (Baird et al., 2003) | TGCTCCGTGCATCTGGCATC | V |
|  | (Baird et al., 2003) |  |  |
|  | XpYp427G/415C | GGTTATCGACCAGGTGCTCC | $66.5{ }^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2003) |  |  |
|  | $\mathrm{XpYp} 427 \mathrm{~A} / 415 \mathrm{~T}$ | GGTTATCAACCAGGTGCTCT | $66.5{ }^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2003) <br> XpYpE2 | TTGTCTCAGGGTCCTAGTG | $63^{\circ} \mathrm{C}$ |
|  | $\begin{aligned} & \text { (Baird et al., 2003) } \\ & \text { XpYpB2 } \end{aligned}$ | TCTGAAAGTGGACCWATCAG | $66^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2003) | C | $65^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2000) |  |  |
|  | 12qnull3 | GATGTCTGAGTGGATTCAGACATG | $67^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2000) | GGCTGAACTATAGCCTCTGC | $63{ }^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2006) |  |  |
|  | TS30T | CTGCTTTTATTCTCTAATCTGCTCCCT | $66.5{ }^{\circ} \mathrm{C}$ |
|  | (Baird et al., 1995) |  |  |
|  | TS30A | GTGCTTTTATTCTCTAATAATCTCCCA | $66.5{ }^{\circ} \mathrm{C}$ |
|  | (Baird et al., 1995) |  |  |
|  | 12q-197A | GGGAGATCCACACCGTAGCA | $66^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2000) |  |  |
|  | 12q-197G | GGGAGATCCACACCGTAGCG | $66^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2000) |  |  |
|  | 12qdeletionSTELA | TGAGCATTCATGAGCATTACAGG | $67^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2000) |  |  |
|  | $12 q B$ | ATTTTCATTGCTGTCTTAGCACTGCAC | $61^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2000) |  |  |
|  | TAG TelX | TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCTNACCCTC | $66^{\circ} \mathrm{C}$ |
|  | (Baird et al., 1995) |  |  |
|  | TAG TelW | TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCTNACCCTA | $66^{\circ} \mathrm{C}$ |
|  | (Baird et al., 1995) |  |  |
|  | TAG TelJ | TCATGCGTCCATGGTCCGGAACCCCAACCCCAACCCCAACCCC | $66^{\circ} \mathrm{C}$ |
|  | (Coleman et al., 1999) <br> TAG TelX-NED | (NED)-TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCTNACCCTC | $66^{\circ} \mathrm{C}$ |
|  | TAG TelW-FAM | (FAM)-TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCTNACCCTA | $66^{\circ} \mathrm{C}$ |
|  | TAG TelJ-HEX | (HEX)-TCATGCGTCCATGGTCCGGAACCCCAACCCCAACCCCAACCCC | $66^{\circ} \mathrm{C}$ |
|  | TAG TelCTA2-FAM | (FAM)-TCATGCGTCCATGGTCCGGACCCTTACCCTTRCCCTARCCCTAG | $66^{\circ} \mathrm{C}$ |
|  | TAG TelY-HEX | (HEX)-TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCTNACCCTG | $66^{\circ} \mathrm{C}$ |
| 号 | TS primer | AATCCGTCGAGCAGAGTT | $59^{\circ} \mathrm{C}$ |
|  | (Kim et al., 1994) |  |  |
|  | RP primer <br> (Kim et al., 1994) | GCGCGG(CTTACC) ${ }_{3}$ CTAACC | $59^{\circ} \mathrm{C}$ |
| ¢ | TS-30ARev | AATCAGCAGGATGTGGGA | $68^{\circ} \mathrm{C}$ |
|  | (Hills, 2004) |  |  |
|  | TS-30TRev | TATCAGCAGGATGTGGGT | $68^{\circ} \mathrm{C}$ |
|  | (Hills, 2004) |  |  |
|  | 17p2 | GAGTCAATGATTCCATTCCTAGC | $60^{\circ} \mathrm{C}$ |
|  | (Baird et al., 2006) |  |  |
|  | pSup1 (Mendez- | CAGAAAGCGAAGGAGCAAAG | $50^{\circ} \mathrm{C}$ |
|  | Bermudez et al., |  |  |
|  | HHV6 probe 31 F | TCGGAAGCGGAGATTTCTAA | $49^{\circ} \mathrm{C}$ |
|  | HHV6 probe 31 R | CCACCTTCGCAACAACAATA | $49^{\circ} \mathrm{C}$ |
|  | HHV6A probe 49 F | CAGGAAAGGGACGGTGATAA | $50^{\circ} \mathrm{C}$ |
|  | HHV6A probe 49 R | ATCGAAAGCACCACCTTCAC | $50^{\circ} \mathrm{C}$ |

## C $\quad$ Multiple alignment for OL A31

The control amplicon for OL A31 was sequenced using the minION ${ }^{\text {TM }}$ device. In the next pages, a multiple alignment generated using ClustalW in the MacVector software, V11.1.1 is shown. The reference sequence obtained by Enjie Zhang using Ion Torrent ${ }^{\text {TM }}$ sequencing is named: "OL A31 Ref"; the consensus generated by IGV from the alignment of the 1D reads is called "1d.OL A31" and for the 2D reads, "1d.OL A31". Each nucleotide is colour coded and the position of each nucleotide is indicated at the top of the alignment in red numbers. Letters are: A: adenine; T: thymine; C: cytosine; G: guanine; R: purine (A or G); Y: pirimidine (T or C); N: any nucleotide; W: Weak (A or T); S: Strong (C or G) and K: Keto (G or T).

1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.


1d.OL A31
2d.OL A31
OL A31 Ref.
consensu


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus




1d.OL A31
2d.OL A31
OL A31 Ref
consensus

| 290 |  |  |  |  |  |  |  |  |  | 300 |  |  |  |  |  |  | 310 |  |  |  |  | 320 |  |  |  |  |  |  |  | 330 |  |  |  |  |  | 340 |  |  |  |  | 350 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| N | $\mathrm{N} N$ | $\mathrm{N} N$ | NN N | $N \mathrm{~N}$ A | A ${ }^{\text {N }}$ | $N \mathrm{~N}$ |  | $N \mathrm{~N}$ | N/N | N |  | N |  | $N \mathrm{~N}$ | $N \mathrm{~N}$ |  | N N | N | N | N | N | N | $N$ | N/N |  | N | N |  | N | N | N | N N | N | $N$ | N N |  | $N \mathrm{~N}$ | N |  | $\mathfrak{r l}$ | A | N | $N$ | N |
|  | G | G | GA | A C A | A |  |  |  | GG | A |  |  | G T |  | G |  |  | A |  |  | G | G | A | A A |  |  | G |  | G |  |  | A A |  | A | A G |  | GA | A |  | G | A |  |  |  |
|  | GT | TG | GA | A C A | A |  | 1 |  | GG | A |  | G | G |  | G |  |  | A |  |  | G | G | A | A A |  | A | G |  | G |  |  | A A |  | A | A $G$ |  | GA | A |  | G | A |  |  |  |
|  | GT |  | G A | A ${ }^{\text {A }}$ | A |  |  |  | G/G | A |  |  | G |  | G |  |  | A |  |  | G | G | TA | A\|A |  |  |  | A |  |  |  | A $A$ |  | A | A/G |  | G A | A |  | G | A |  | A $A$ |  |



1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus



990
1000
1010
1020
1030
1040
1050
1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus



1130

1d.OL A31
2d.OL A31
OL A31 Ref.
consensus

1d.OL A31
2d.OL A31
OL A31 Ref.
consensus

1d.OL A31
2d.OL A31
OL A31 Ref.
consensus
d.OL A31

2d.OL A31
OL A31 Ref.
consensus

| NN |  |  | NNN | NN |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | NN |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| GNA | AGT | G | AAA | A | A | GA | GG | $A^{-1}$ | T TTG | GTG |  | GA | N | CTG |  | GA |  |  | AG |  | GA |  |  | GG |  |  |
| A | AGG | G | A AAA | A | A | GA | GG | A | G | GTG |  | GA | G | G |  | G/A | TTA | A | AG | A | , A |  |  | GG |  |  |
| 6 A | A\|G] | G | A $A_{A} \mid$ A | A | A | GA | GG | A |  |  |  | AGA ${ }^{\text {a }}$ |  |  |  | GA |  |  | A/G |  |  |  |  | GG |  |  |





1340
1350
1360
1370
1380
1390 1400




A A A
AIA
1410 1420 $1430 \quad 1440 \quad 1450 \quad 14601470$

2d.OL A31 L A31 Ref.
consensus

 2d.OL A31

OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus

d.OL A31

2d.OL A31
OL A31 Ref.
consensus


176017701780180018010180

2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus

d.OL A31

2d.OL A31
OL A31 Ref.
consensus


| 2110 | 2120 | 2130 | 2140 | 2150 |
| :---: | :---: | :---: | :---: | :---: |

d.OL

2d.OL A31
OL A31 Ref.
consensus


2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


2390
2400
2410
2420
2430
2440
2450
1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


2d.OL A31
OL A31 Ref.
consensus


2600


2670
2680
2690
2700
2710
2720
2730

2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus

| 28102820 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 2830 |  |  |  |  |  | 2840 |  |  |  |  |  | 2850 |  |  |  |  |  | 2860 |  |  |  |  | 2870 |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | N ${ }^{\text {N }}$ | N/N | N/N\|N | N N | N N | N N N | $\mathrm{N} / \mathrm{N}$ | $\mathrm{N} / \mathrm{N}$ | N | N N | N | N N | N | $\mathrm{N} \times$ | $\mathrm{N} N$ | $N$ | N N | N | N | $\mathrm{N} N$ | N | N | NG/N | N $N$ | $N \sim$ | N | N | N | G | N | N N | N/N | $N$ | $N$ | G | N $N$ | N |  | N $N$ | N | G N | N N N N |
|  |  | A $T$ | T A G | NT | T G | G N | $\mathrm{N} N$ | A | A | N A |  |  | G | G | G |  |  | G |  |  |  | A | A G A | A | A | G |  |  | G | A | G |  | A | A A | G |  | A |  | G |  | G | $N \mathrm{~A}$ |
| T 1 | T T | A T | T A G |  | -G/G | G |  | A | A | A | A |  | G | G | G |  |  | G |  |  | T |  | A GA |  | A | G |  |  | G | A | G |  | A | A A | G | 1 | A |  | G |  | G | A |
| 1 | T T | A T | TIA\|G |  | TGIG | G |  | A | A | A | A |  | G T | G | G |  |  | G |  |  | T |  | A\|G|A | A T | A | G |  |  | G] | A | G |  | A | A $A$ | G | T | A |  | G |  | G | A |
|  |  |  |  |  | 880 |  |  |  |  |  | 890 |  |  |  |  |  | 290 |  |  |  |  |  |  | 291 |  |  |  |  |  |  | 220 |  |  |  |  | 29 |  |  |  |  |  | 2940 |

1d.OL A31
2d.OL A31
OL A31 Ref.




3020
3030
3040
3050
3060
3070
3080

2d.OL A31
OL A31 Ref.
consensus
d.OL A31

2d.OL A31
OL A31 Ref
consensus


| 3160 | 3170 | 3180 | 3190 | 3200 |
| :--- | :--- | :--- | :--- | :--- |
| 3210 | 3220 |  |  |  |

2d.OL A31
OL A31 Ref.

| N $\mathrm{N} / \mathrm{N}$ | NN/N | NN/N | $N \mathrm{~N}$ | N N | N | NG | N | $\mathrm{N} / \mathrm{N}$ | G | N N | N N | N | N | $\mathrm{N} / \mathrm{N}$ | N | NA | N | N | $N$ | $N$ | N | $\mathrm{N} N$ | G | $N$ | N | N | N | N | N |  | $N$ | N | G | $N$ | N | $N$ | N | N |  | N |  | $\mathrm{N} N$ | N | $N$ | N | N |  | $N$ | $\mathrm{N} N$ | N |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| A | A $N$ |  | A | A | G | A G | A |  | G |  | N | G | A | A |  | A |  | T |  |  | T |  |  | A |  |  | G | G |  |  |  |  | G | A |  |  | A | A |  |  |  | N N |  |  |  | A |  |  |  | G |
| A | A |  | A | A |  | A G | A |  | G T |  | TT |  | A |  | T | A |  | T | T |  | T |  |  | A |  |  | G | G |  |  |  |  | G | A |  | A | A | A |  | T |  |  |  |  |  | A |  |  |  | G |
| A | A |  | A | A |  | A G | A |  | G T |  |  |  | A |  |  | A |  |  |  |  |  |  |  | A |  |  | G | G |  |  |  |  | G | A |  | A | A | A |  |  |  |  |  |  |  | A |  |  |  | G |



2d.OL A31
OL A31 Ref.

2d.OL A31
OL A31 Ref.
consensus


3440
3450
3460
3470
3480
3490
3500


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus



36503660
3670
3680
3690
3700
3710
1d.OL A31
2d.OL A31
OL A31 Ref.
consensus




3790
3800
3810
3820
3830
3840 3850

d.OL

2d.OL A31
OL A31 Ref.
consensus



1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


1d.OL A31
2d.OL A31
OL A31 Ref.
consensus




1d.OL A31
2d.OL A31
OL A31 Ref.
consensus


## D Multiple alignment for KUK A49

The control amplicon for KUK A49 was sequenced using the minION ${ }^{\text {TM }}$ device. In the next pages, a multiple alignment generated using ClustalW in the MacVector software, V11.1.1 is shown. The reference sequence obtained by Enjie Zhang using Ion Torrent ${ }^{\text {TM }}$ sequencing is named: "KUK49_final "; the consensus generated by IGV from the alignment of the 1D reads is called "kuk.1d" and for the 2 D reads, "kuk.2d". Each nucleotide is colour coded and the position of each nucleotide is indicated at the top of the alignment in red numbers. Letters are: A: adenine; T: thymine; C: cytosine; G: guanine; R: purine (A or G); Y: pirimidine (T or C); N: any nucleotide; W: Weak (A or T); S: Strong (C or G) and K: Keto (G or T).






|  |  | 1760 | 1770 | 1780 |  | 1790 |  | 1800 |  | 1810 |  |  | 1820 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | kuk.2d | AGGAGATTTGAT ${ }^{\text {a }}$ GG | GAAAIAA A AlG | A $A$ A $A$ C G ${ }^{\text {a }}$ | A ${ }^{\text {G }}$ | GA/G | GGA | GA |  | ATTGAG | IGGAG |  | GGGA |
|  | kuk.1d | AGAAGATTGATGG | GAAAAA AG | A AAA GT | AG | GAGT | GGA | GA |  | ATGAG | GGAG |  | GGGA |
|  | KUK49_final | AGAAGATTGAT GG | GAAAAA AG | AAAA GT | A ${ }^{\text {G }}$ | GAGGT | GGA | GA |  | ATTGAA | GGGAG |  | GGGA |
|  | consensus |  |  |  | A\|G | G $\mathrm{A}_{\mathrm{G}} \mathrm{GT}$ | G/GA | GA |  | ATTGAG | AGIGAlG |  | GGG ${ }^{\text {a }}$ |
|  |  | 1830 | 1840 | 1850 |  | 1860 |  | 1870 |  | 1880 |  |  | 1890 |
|  | kuk.2d | ATATTCAAIGCG | $\mathrm{GCl\mid} A\|A\| A$ | GITGA GG | [ ${ }^{\text {A }}$ A ${ }^{\text {a }}$ | A A A G | GAA | A ${ }^{\text {G G }}$ G | A $A$ A | GATA | C 6 |  | GA/G |
|  | kuk.1d | ATATTCAAGGG | GClCAAA | GTTGTGAC GG | AAG | A A ${ }^{\text {a }}$ | GAA | A ${ }^{\text {G G }}$ | A AA | GATA | A 6 | G | GAG |
|  | KUK49_final | ATATTAAAGGG | GC C A A A A | GTTGTGGA GG | A A/G | A A AG | GAA | A AGG | A AAA | GATA | AC $G$ |  | GAG |
|  | consensus |  |  | ITGTGGAEGG | A $\\|_{A} \mid G$ | A\|A] | $\underline{G\|A\| A} \mid$ | A $A$ G $G$ G |  | G/ATIA | A ${ }^{\text {G }}$ | G | G $\mathrm{A}_{\mathrm{G} / \mathrm{T}}$ |
|  |  | 1900 | 1910 | 1920 |  | 1930 |  | 1940 |  | 1950 |  |  | 1960 |
|  | kuk.2d |  | GTA ATA G | TAAAGCGITTA |  |  |  | GA A A A |  |  | TGA AT |  |  |
|  | kuk.1d | AAAGGTGAGAG | GTATAG | AAAGCGTTA |  | GG | ATAA | , |  | AAAAGAAG | G AAT |  | GG |
| 暑 | KUK49_final | GAAAGGTGAAGAG | GTATAGTI | TAAGG GTTAAT |  | GGc GGc | ATTAA | G/ata |  | AAAAGAAAG | G $\mathrm{A}_{\text {a }}$ |  | GG |
|  | consensus |  | GITAATA\|GIT |  | GGGA | GGIGGI | AITAA | IGA $\mathrm{A}^{\text {a }}$ A | A $A$ A $/ 9$ | $[A\|A\| A\|G\| A\|A\| C$ | AGA $\mathrm{A}_{\text {a }}$ |  | G |
|  |  | 1970 | 1980 | 1990 |  | 2000 |  | 2010 |  | 2020 |  |  | 2030 |
|  | kuk.2d | AGGTACICTGGTAG | GGTTTTA | $G^{\text {G A A G }}$ |  | A A A A |  | G A\|A |  |  | A A G ${ }^{\text {I }}$ |  | A A A A |
|  | kuk.1d | AGTACCTGTAG | GGTTTA | G AC G | AGG | ATTAAAA |  | $G{ }^{\text {a }}$ |  | G | AAGT |  | $A A A A$ |
|  | KUK49_final | AGTACICTGTAG | GGTTTA | G A G ${ }^{\text {a }}$ | AGG | ATTTAAAA | GA | GG A AA | A AA | GTClC G | AAGIT |  | $A\|A\| A \mid$ |
|  | consensus | AIGTAACICTTGITAG | Gclagatitia | GGAACGTT | AGGG | ATTTAAAA | TGA | GGICA $A^{\prime}$ A | A $A^{\text {a }}$ AT | GITCICGIA |  | A | $A\|A\| A \mid A T T$ |
|  |  | 2040 | 2050 | 2060 |  | 2070 |  | 2080 |  | 2090 |  |  | 2100 |
|  | kuk.2d | ATTCICTGGG GATTG | AGIATTG | TGTAGGGGAT |  | A ${ }^{\text {G }}$ | Gc G | GTGGA |  | GAIAIGA |  | A | GG/AG |
|  | kuk.1d | ATCCCTGGGGA | AGATG | TGTAGGGGAT | AGA | AG | GCG | GTGGA | AG | GAAGA |  | A | GGAG |
|  | KUK49_final | ATCCCTGGG GAT G | GAGATG | TGTAGGGGATG | GTAGA | TAGTTT | Gc G | GTGGA | AG | GAAGAI |  | A | GGAG |
|  | consensus | AITCICITGGGGGATTG | GAGAATG | TGITAGGGGATTG | GTAGGA | TA GTTTT | GIG | TGTGGG | AGIT |  | +1. | A | $\underline{G G / A / G}$ |









## E Multiple alignment for XpYp telomere

The A allele of the XpYp telomere from the LoVo cell line ( $\mathrm{MSH}^{-/-}$) was sequenced using the minION ${ }^{\text {TM }}$ device. Two libraries (one containing both XpYp alleles, and another containing only A allele) generated 1D and 2D reads that were first aligned to the flanking region of the A allele to filter the reads (as explained in section 4.3.3). Later, the filtered reads were aligned to a telomere consensus generated based on the TVR maps (described in section 4.9) called "Telomere.reference" for ClustalW multiple alignment. The consensus generated by IGV from the alignment of the 1D and 2D reads for both alleles are called: "1d.E2.filtered" and "1d.E2.filtered" respectively. The consensus generated by IGV from the alignment of 1 D and 2 D reads for A allele are called: "1d.hapA.filtered" and "1d.hapA.filtered" respectively. Each nucleotide is colour coded and the position of each nucleotide is indicated at the top of the alignment in red numbers. Letters are: A: adenine; T: thymine; C: cytosine; G: guanine; R: purine (A or G); Y: pirimidine (T or C); N: any nucleotide; W: Weak (A or T); S: Strong (C or G) and K: Keto (G or T).



|  |  |  |  | 570 |  | 80 |  |  | 590 |  |  |  | 600 | 00 |  |  | 610 |  |  | 620 |  |  | 630 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1d.E2.filtered | $G$ | A GGG ${ }^{\text {a }}$ | A GGGGT | A GGG |  | TAGGG |  | A GIGG |  |  | GGG |  | A GGG |  | AGG | G | TAGGG |  | A GGG |  | R G GGT | $R$ |
|  | 2d.E2.filtered | $G$ | AGGGT | AGGGT | AGGG |  | TAGGG |  | AGGG |  |  | GGG |  | AGGGT |  | AGG | G | AGGG |  | A GGG |  | R GGGT | R |
|  | 1d.hapA.filtered | $G$ | AGGGT | AGGGT | A GGG |  | TAGGG |  | A GGG |  |  | GGG |  | A GGGT |  | AGG | G | TAGGG |  | A GGG |  | RGGG | R |
|  | 2d.hapA.filtered | $G$ | AGGGT | AGGGT | AGGG |  | TAGGG |  | A GGG |  |  | GGG |  | AGGGT |  | AGG |  | AGGG |  | AGGG |  | R GGGT | R |
|  | Telomere.reference | G | AGGGGT | AGGGT | A GGGG |  | TAGGG |  | AGGGG |  |  | GGG |  | AGGGT |  | AGG |  | AGGG |  | A GGG |  | GG/GGT | 6 |
|  | consensus |  | AGGGGT | AGGGGT | A/GGG |  | TAGGGG |  | A/GGG |  |  | GGG] |  | AGGGGT |  | AGGG |  | A G/G] |  | A/G]GG |  | GGGGGT | TG |
|  |  |  |  | 640 |  | 50 |  |  | 660 |  |  |  | 670 | 70 |  |  | 680 |  |  | 690 |  |  | 700 |
|  | 1d.E2.filtered | NG | RGG | GGTTRGG |  | GGG |  | GGG |  |  |  |  |  | RGG | GG |  | GGG | GTTRG | GGT |  |  | R GG | GG |
|  | 2d.E2.filtered | NG | GTTRGG | GGTTRGG | GTTRG | GGG | GTTRG | GGG | R | GG |  | RG | GGT | RGG | GG |  | GGG | GTTRGG | GGT |  | GGT | RGG | GG |
|  | 1d.hapA.filtered | $\cdots$ | GTTRGG | GGTTRGG | GTTRG | GGG | GTTRG | GGG | R |  |  |  | GG7 | TTRGG |  |  | GGG | GTTRGG | GGT |  | GGT | RGG | GG |
|  | 2d.hapA.filtered | GG | GTTRGG | GGTTRGG | GTTRG | GGG | GTTRG |  | R |  |  | NG | GG7 | TTNGG |  |  | GGG | GTTRGG |  |  | GGT | RGG | GG |
|  | Telomere.reference | GG | GTT GGG | GGTT G GG | GTT GG | GGG |  |  |  |  |  |  |  | TTGGG |  | G | GG/G | GTTGGG |  | GG | GGT | G/GG |  |
|  | consensus |  | TTGGGG | GGTT TGGG | GTTTGG | GGG | GTTTGG |  | GG |  |  |  |  | TTTGGG |  | TTTG | GGGG | GTTTGGG |  |  |  | $\mathrm{G} / \mathrm{G} / \mathrm{G}$ | GG] |
| L |  |  |  | 710 |  | 20 |  |  | 730 |  |  |  | 740 | 40 |  |  | 750 |  |  | 760 |  |  | 770 |
| $\underset{\sim}{\Delta}$ | 1d.E2.filtered |  | GGG | R GGGG | GGG |  | GGG |  | GG |  | GGG |  | A G $G$ | GG |  | GG |  | GGG | NGG | $G G$ | RGG | GG | R\|GG |
| $<$ | 2d.E2.filtered |  | GGG | RGGGTTA | GGG |  | GGG |  | GG |  | GGG |  | AGG | GGTTR |  | GG |  | GGG | NGG | GG | RGG | GG | RGG |
|  | 1d.hapA.filtered |  | GGG | RGGGTTA | GGG |  | GGG |  | GG7 |  | GGG |  | AGG | GGTTR |  | GG |  | GGG | NGG | GG | RGG | GG | RGG |
|  | 2d.hapA.filtered |  | GGG | RGGGTTA | $A G G G$ | $\mathrm{TAG}$ | $\mathrm{GGG7}$ |  |  |  |  |  | A GG | GGTTR |  | GG |  | GGG | RGG | GG | RGG | GG | RGG |
|  | Telomere.reference |  | $\mathrm{G} \mid \mathrm{G} G$ | GGGGTTA | $G G G G$ | $\mathrm{TAGG}$ | $\mathrm{G} \mid \mathrm{G}$ |  | $i G \mathrm{GT}$ |  |  |  | AGGG | GGTTG |  |  |  |  |  |  |  | GG | GGG |
|  | consensus |  | G GGGTTTG | GGGGGTTTA | G/G\|G] |  | GG\|GT | TAG | G\|GT |  | GIGG |  |  | G GGTTTG | GGG | GGG] |  | G GGGT | R\|G] | GGT | G/G] | G GTTTG | G/GG |
|  |  |  |  | 780 |  | 90 |  |  | 800 |  |  |  | 810 | 10 |  |  | 820 |  |  | 830 |  |  | 840 |
|  | 1d.E2.filtered | $G$ | R GGGGT | R G GGT | A GGG |  | TAGGG |  | A GIGG |  |  | GGG |  | A GGGG |  | AGG |  | AGGG |  | A GGG |  | A G G G | A |
|  | 2d.E2.filtered | $G$ | R GGGT | R GGGT | A GGG |  | TAGGG |  | AGGG |  |  | GGG |  | AGGG |  | AGG |  | AGGG |  | A GGG |  | AGGGT | A |
|  | 1d.hapA.filtered | $G$ | R GGGT | RGGGT | A GGG |  | TAGGG |  | A GGG |  |  | GGG |  | AGGG |  | AGG |  | AGGG |  | AGGG |  | A GGGT | A |
|  | 2d.hapA.filtered | GT | RGGGT | RGGGT | A GGG |  | TAGGG |  | AGGG |  |  | GGG |  | AGGGT |  | AGG | G | AGGG |  | AGGG |  | AGGG | A |
|  | Telomere.reference | GT | TGGGGI | GGGGT | AGGG |  | TAGGG |  | AGGG |  |  | GGG |  | AGGGT |  | AGG |  | AGGG |  | A)GGG |  | AGGGI | A |
|  | consensus |  | TGGGGGT | TGGGGGT | A/GGG | G | TAGGGG |  | A/GGG |  | AIG | GGG |  | A GGGGT |  | AGGG | G | TAGGGG |  | A/GGG |  | AGGGG] | A |

















## F Students' contribution to the families study

The table below shows the contribution of each student to the project. Abbreviations are: TEL: code for control families; LRI: code for LS samples from Leicester's Royal Infirmary; BWH: code for LS samples from Birmingham Women's Hospital; FT: Frances Tippins; CG: Carmen Garrido.

Recruitment refers to: ethical approval (for BWH families recruited by CG, documents generated by FT were used. For TEL families, new documents were generated), family selection based on pedigree, family contact (for BWH families recruited by CG, this was done by J. Hoffman in BWH), saliva kit preparation (including documents) and shipment. Sample processing refers to: aliquoting, stock storage, DNA extraction and STELA analysis (from PCR to telomere sizing). Analysis refers to: statistical analysis.

| Family | Recruitment | Sample processing | Analysis |
| :--- | :---: | :---: | :---: |
| TEL F1 | CG | CG | CG |
| TEL F2 | CG | CG | CG |
| TEL F3 | CG | CG | CG |
| TEL F4 | CG | CG | CG |
| TEL F6 | CG | CG | CG |
| TEL F7 | CG | CG | CG |
| TEL F8 | CG | CG | CG |
| TEL F9 | CG | CG | CG |
| TEL F11 | CG | CG | CG |
| TEL F12 | CG | CG | CG |
| LRI F1 | FT | FT | CG |
| LRI F4 | FT | CG | CG |
| LRI F5 | FT | FT | CG |
| LRI F7 | FT | FT | CG |
| LRI F12 | FT | CG | CG |
| LRI F14 | FT | CG | CG |
| BWH F1a | CG | CG | CG |
| BWH F3a | CG | CG | CG |
| BWH F5a | CG | CG | CG |
| BWH F6a | CG | CG | CG |
| BWH F11a | CG | CG | CG |
| BWH F2 | FT | FT | CG |
| BWH F42 | FT | FT | CG |
| BWH F55 | FT | CG | CGH F106 |

## G Technical reproducibility for control families

The scatter plots in $\mathbf{A}$ show telomere length assayed by single telomere length analysis (STELA) using the XpYpE 2 primer for the control DNA extracted from the lymphoblastoid cell line KK and run in every gel to assess technical reproducibility. The legend in the X -axis refers to the family and the corresponding gel number. $\mathbf{B}$ shows the median and interquartile for A and Kruskal Wallis test showed no differences between any gel.


# H Technical reproducibility for Lynch syndrome families 

The same control DNA than in appendix G was used to assess technical reproducibility in STELA gels for the Lynch syndrome families. A and C show medians and interquartile ranges for $12 q$ and XpYp gels respectively. Light green scatter plots show data generated by Frances Tippins (FT) and dark green plots show data generated by Carmen Garrido (CG). B and $\mathbf{D}$ show the mean values for A and C respectively. Purple colour represents the gels that were excluded as the control was significantly different by Kruskal-Wallis ranking test.

A


B


C


D


## I Control families pedigrees

The next five pages contain two control families per page. For each family, the code is shown in bold letters after the prefix "TEL" next to the pedigree. Males are indicated with squares and females with circles. The code for each individual is shown under their respective symbol preceded by the "TEL + family code" prefix. A "G" in the code indicates generation and a "P", position in the pedigree. The age of each individual in shown below the code in brackets. Below each pedigree, scatter plots showing medians and interquartile ranges (IQ) for 12q (cyan) and XpYp (purple) telomeres measured by single telomere length analysis (STELA) is shown. Details of the medians and IQ are shown in the table below each scatter plot.




TEL F7


TEL F7 G2P1 (31.4)

TEL F8

TEL F9

TEL F9 G2P3 TEL F9 G2P2 TEL F9 G2P1



TEL F12


TEL F12 G2P3
TEL F12 G2P2


## J Lynch syndrome families pedigrees

The next twelve pages contain two Lynch syndrome (LS) families per page. For each family, the code is shown in bold letters after the prefix "LRI" for families collected from Leicester Royal Infirmary or "BWH" for families collected form Birmingham Women's Hospital next to the pedigree. Males are indicated with squares and females with circles. The code for each individual is shown under their respective symbol preceded by the "LRI or BWH + family code" prefix. A "G" in the code indicates generation and a "P", position in the pedigree. The age of each individual in shown below the code in brackets. In addition, the genotype for the mismatch repair gene (either MSH2 or MLH1) is shown below the age and "WT" refers to the wild types, non-mutation carriers. Furthermore, cancer-affected patients are shown with coloured symbols and the age of onset is also indicated after "@". Below each pedigree, scatter plots showing medians and interquartile ranges (IQ) for $12 q$ (cyan) and XpYp (purple) telomeres measured by single telomere length analysis (STELA) is shown. Details of the medians and IQ are shown in the table below each scatter plot. Abbreviations for the tumours and colour code is: purple for colorectal cancer (CRC), green for endometrial cancer (EC), orange for urothelial carcinoma (UBC), blue for malignant melanoma (MM) and red for basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and thyroid cancer (TC).


LXII



LRI F14


LXIV



LXVI

BWH F42

BWH F55


LXVIII


LXIX




LXXII


## Bibliography

Acharya, S., Foster, P. L., Brooks, P., \& Fishel, R. (2003). The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. Molecular cell, 12(1), 233-46. URL http://www.ncbi.nlm.nih.gov/pu.bmed/12887908.

Achi, M. V., Ravindranath, N., \& Dym, M. (2000). Telomere length in male germ cells is inversely correlated with telomerase activity. Biology of reproduction, 63(2), 591-8. URL http://www.ncbi.nlm.nih.gov/pubmed/10906069.

Ahmed, D., Eide, P. W., Eilertsen, I. A., Danielsen, S. A., Eknæs, M., Hektoen, M., Lind, G. E., \& Lothe, R. A. (2013). Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis, 2(0424), e71. URL http://www.ncbi.nlm.nih.gov/pubmed/24042735.

Ahuja, N., Mohan, a. L., Li, Q., Stolker, J. M., Herman, J. G., Hamilton, S. R., Baylin, S. B., \& Issa, J. P. (1997). Association between CpG island methylation and microsatellite instability in colorectal cancer. Cancer research, 57(16), 3370-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/9269998.

Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P., \& Griffith, J. D. (1997). MutS mediates heteroduplex loop formation by a translocation mechanism. The EMBO journal, 16(14), 4467-76. URL http: //www.ncbi.nlm.nih.gov/pubmed/9250691.

Allsopp, R. C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E. V., Futcher, a. B., Greider, C. W., \& Harley, C. B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proceedings of the National Academy of Sciences of the United States of America, 89(21), 10114-8. URL http://www.ncbi.nlm.nih.gov/pubmed/1438199.

Alter, B. P., Baerlocher, G. M., Savage, S. a., Chanock, S. J., Weksler, B. B., Willner, J. P., Peters, J. a., Giri, N., \& Lansdorp, P. M. (2007). Very short telomere length by flow fluorescence in situ hybridization identifies patients with dyskeratosis congenita. Blood, 110(5), 1439-47. URL http://www.ncbi.nlm.nih.gov/pubmed/17468339.

Alter, B. P., Rosenberg, P. S., Giri, N., Baerlocher, G. M., Lansdorp, P. M., \& Savage, S. a. (2012). Telomere length is associated with disease severity and declines with age in dyskeratosis congenita. Haematologica, 97(3), 353-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/22058220.

Althubiti, M., Lezina, L., Carrera, S., Jukes-Jones, R., Giblett, S. M., Antonov, A., Barlev, N., Saldanha, G. S., Pritchard, C. a., Cain, K., \& Macip, S. (2014). Characterization of novel markers of senescence and their prognostic potential in cancer. Cell death \& disease, 5(11), e1528. URL http://www.ncbi.nlm.nih.gov/pubmed/25412306.

Andrew, T., Aviv, A., Falchi, M., Surdulescu, G. L., Gardner, J. P., Lu, X., Kimura, M., Kato, B. S., Valdes, A. M., \& Spector, T. D. (2006). Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. American journal of human genetics, 78(3), 480-6. URL http: //www.ncbi.nlm.nih.gov/pubmed/16400618.

Arnoult, N. \& Karlseder, J. (2015). Complex interactions between the DNA-damage response and mammalian telomeres. Nature structural \& molecular biology, 22(11), 859-66. URL http://www.ncbi.nlm.nih.gov/pubmed/26581520.

Aston, K. I., Hunt, S. C., Susser, E., Kimura, M., Factor-Litvak, P., Carrell, D., \& Aviv, A. (2012). Divergence of sperm and leukocyte age-dependent telomere dynamics: implications for male-driven evolution of telomere length in humans. Molecular human reproduction, 18(11), 517-22. URL http://www.ncbi.nlm.nih.gov/pubmed/22782639.

ATCC (2014). Animal Cell Culture Guide. Tech. rep., ATCC. URL www. atcc.org.
Aubert, G., Baerlocher, G. M., Vulto, I., Poon, S. S., \& Lansdorp, P. M. (2012a). Collapse of telomere homeostasis in hematopoietic cells caused by heterozygous mutations in telomerase genes. PLoS genetics, 8(5), e1002696. URL http: //www.ncbi.nlm.nih.gov/pubmed/22661914.

Aubert, G., Hills, M., \& Lansdorp, P. M. (2012b). Telomere length measurement-caveats and a critical assessment of the available technologies and tools. Mutation research, 730(1-2), 59-67. URL http://www.ncbi.nlm.nih.gov/pubmed/21663926.

Aviv, A., Chen, W., Gardner, J. P., Kimura, M., Brimacombe, M., Cao, X., Srinivasan, S. R., \& Berenson, G. S. (2009). Leukocyte telomere dynamics: longitudinal findings among young adults in the Bogalusa Heart Study. American journal of epidemiology, 169(3), 323-9. URL http://www.ncbi.nlm.nih.gov/pubmed/19056834.

Aviv, A., Valdes, A., Gardner, J. P., Swaminathan, R., Kimura, M., \& Spector, T. D. (2006). Menopause modifies the association of leukocyte telomere length with insulin resistance and inflammation. The Journal of clinical endocrinology and metabolism, 91(2), 635-40. URL http://www.ncbi.nlm.nih.gov/pubmed/16303830.

Baerlocher, G. M., Vulto, I., de Jong, G., \& Lansdorp, P. M. (2006). Flow cytometry and FISH to measure the average length of telomeres (flow FISH). Nature protocols, 1(5), 2365-76. URL http://www.ncbi.nlm.nih.gov/pubmed/17406480.

Baird, D. M. (2008). Telomere dynamics in human cells. Biochimie, 90(1), 116-21. URL http://www.ncbi.nlm.nih.gov/pubmed/17854970.

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(1), 45-51. URL http://www.ncbi.nlm.nih.gov/pubmed/16311252.

Baird, D. M., Coleman, J., Rosser, Z. H., \& Royle, N. J. (2000). High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: implications for telomere biology and human evolution. American journal of human genetics, 66(1), 235-50. URL http://www.ncbi.nlm.nih.gov/pubmed/10631154.

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 $\mathrm{Xp} / \mathrm{Yp}$ telomere. The EMBO journal, 14(21), 5433-43. URL http: //www.ncbi.nlm.nih.gov/pubmed/7489732.

Baird, D. M., Rowson, J., Wynford-Thomas, D., \& Kipling, D. (2003). Extensive allelic variation and ultrashort telomeres in senescent human cells. Nature genetics, 33(2), 203-7. URL http://www.ncbi.nlm.nih.gov/pubmed/12539050.

Bak, S. T., Sakellariou, D., \& Pena-Diaz, J. (2014). The dual nature of mismatch repair as antimutator and mutator: for better or for worse. Frontiers in genetics, 5(August), 287. URL http://www.ncbi.nlm.nih.gov/pubmed/25191341.

Baker, S. M., Bronner, C. E., Zhang, L., Plug, a. W., Robatzek, M., Warren, G., Elliott, E. a., Yu, J., Ashley, T., Arnheim, N., Flavell, R. a., \& Liskay, R. M. (1995). Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. Cell, 82(2), 309-19. URL http: //www.ncbi.nlm.nih.gov/pubmed/7628019.

Balganesh, T. S. \& Lacks, S. a. (1985). Heteroduplex DNA mismatch repair system of Streptococcus pneumoniae: cloning and expression of the hexA gene. Journal of bacteriology, 162(3), 979-84. URL http://www.ncbi.nlm.nih.gov/pubmed/3997786.

Balin, A. K., Reimer, R. J., Reenstra, W. R., Lilie, S. M., Leong, I., Sullivan, K., \& Allen, R. G. (2010). Effects of oxygen, growth state, and senescence on the antioxidant responses of WI-38 fibroblasts. Age (Dordrecht, Netherlands), 32(4), 435-49. URL http: //www.ncbi.nlm.nih.gov/pubmed/20473639.

Ban, C., Junop, M., \& Yang, W. (1999). Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. Cell, 97(1), 85-97. URL http: //www.ncbi.nlm.nih.gov/pubmed/10199405.

Ban, C. \& Yang, W. (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. Cell, 95(4), 541-52. URL http: //www.ncbi.nlm.nih.gov/pubmed/9827806.

Bandaria, J. N., Qin, P., Berk, V., Chu, S., \& Yildiz, A. (2016). Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin. Cell, 164(4), 735-46. URL http: //www.ncbi.nlm.nih.gov/pubmed/26871633.

Barma, D. K., Elayadi, A., Falck, J. R., \& Corey, D. R. (2003). Inhibition of telomerase by BIBR 1532 and related analogues. Bioorganic \& medicinal chemistry letters, 13(7), 1333-6. URL http://www.ncbi.nlm.nih.gov/pubmed/12657276.

Barwell, J., Pangon, L., Georgiou, A., Docherty, Z., Kesterton, I., Ball, J., Camplejohn, R., Berg, J., Aviv, A., Gardner, J., Kato, B. S., Carter, N., Paximadas, D., Spector, T. D., \& Hodgson, S. (2007). Is telomere length in peripheral blood lymphocytes correlated with cancer susceptibility or radiosensitivity? British journal of cancer, 97(12), 1696-700. URL http://www.ncbi.nlm.nih.gov/pubmed/18000505.

Bataille, V., Kato, B. S., Falchi, M., Gardner, J., Kimura, M., Lens, M., Perks, U., Valdes, A. M., Bennett, D. C., Aviv, A., \& Spector, T. D. (2007). Nevus size and number are associated with telomere length and represent potential markers of a decreased senescence in vivo. Cancer epidemiology, biomarkers \& prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 16(7), 1499-502. URL http://www.ncbi.nlm.nih.gov/pubmed/17627017.

Baumann, P. \& Cech, T. R. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. Science (New York, N.Y.), 292(5519), 1171-5. URL http: //www.ncbi.nlm.nih.gov/pubmed/11349150.

Beal, M. A., Rowan-Carroll, A., Campbell, C., Williams, A., Somers, C. M., Marchetti, F., \& Yauk, C. L. (2015). Single-molecule PCR analysis of an unstable microsatellite for detecting mutations in sperm of mice exposed to chemical mutagens. Mutation research, 775, 26-32. URL http://www.ncbi.nlm.nih.gov/pubmed/25863182.

Bechter, O. E., Zou, Y., Walker, W., Wright, W. E., \& Shay, J. W. (2004). Telomeric recombination in mismatch repair deficient human colon cancer cells after telomerase inhibition. Cancer research, 64(10), 3444-51. URL http: //www.ncbi.nlm.nih.gov/pubmed/15150096.

Benetos, a., Okuda, K., Lajemi, M., Kimura, M., Thomas, F., Skurnick, J., Labat, C., Bean, K., \& Aviv, A. (2001). Telomere length as an indicator of biological aging: the gender effect and relation with pulse pressure and pulse wave velocity. Hypertension, 37(2 Pt 2), 381-5. URL http://www.ncbi.nlm.nih.gov/pubmed/11230304.

Benitez-Buelga, C., Vaclová, T., Ferreira, S., Urioste, M., Inglada-Perez, L., Soberón, N., Blasco, M. A., Osorio, A., \& Benitez, J. (2016). Molecular insights into the OGG1 gene, a cancer risk modifier in BRCA1 and BRCA2 mutations carriers. Oncotarget. URL http://www.ncbi.nlm.nih.gov/pubmed/27015555.

Bermudez, A. M. (2007). Telomere stability in cancer cells and the role of DNA mismatch repair. Ph.D. thesis, Leicester.

Bertorelle, R., Rampazzo, E., Pucciarelli, S., Nitti, D., \& De Rossi, A. (2014). Telomeres, telomerase and colorectal cancer. World journal of gastroenterology, 20(8), 1940-50. URL http://www.ncbi.nlm.nih.gov/pubmed/24616570.

Betts, D. H., Perrault, S. D., \& King, W. A. (2008). Low oxygen delays fibroblast senescence despite shorter telomeres. Biogerontology, 9(1), 19-31. URL http: //www.ncbi.nlm.nih.gov/pubmed/17952625.

Bhayadia, R., Schmidt, B. M. W., Melk, A., \& Hömme, M. (2016). SenescenceInduced Oxidative Stress Causes Endothelial Dysfunction. The journals of gerontology. Series A, Biological sciences and medical sciences, 71(2), 161-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/25735595.

Bilaud, T., Koering, C. E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S. M., \& Gilson, E. (1996). The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. Nucleic acids research, 24(7), 1294-303. URL http://www.ncbi.nlm.nih.gov/pubmed/8614633.

Blackburn, E. H. (1991). Structure and function of telomeres. Nature, 350(6319), 569-73. URL http://www.ncbi.nlm.nih.gov/pubmed/1708110.

Blackburn, E. H., Greider, C. W., Henderson, E., Lee, M. S., Shampay, J., \& Shippen-Lentz, D. (1989). Recognition and elongation of telomeres by telomerase. Genome / National Research Council Canada = Génome / Conseil national de recherches Canada, 31(2), 55360. URL http://www.ncbi.nlm.nih.gov/pubmed/2698831.

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 (New York, N.Y.), 279(5349), 349-52. URL http://www.ncbi.nlm.nih.gov/pubmed/9454332.

Boland, C. R. \& Lynch, H. T. (2013). The history of Lynch syndrome. Familial cancer, 12(2), 145-57. URL http://www.ncbi.nlm.nih.gov/pubmed/23546821.

Bonetti, D., Martina, M., Falcettoni, M., \& Longhese, M. P. (2013). Telomereend processing: mechanisms and regulation. Chromosoma, 57-66. URL http: //www.ncbi.nlm.nih.gov/pubmed/24122006.

Boonstra, P. S., Gruber, S. B., Raymond, V. M., Huang, S.-C., Timshel, S., Nilbert, M., \& Mukherjee, B. (2010). A review of statistical methods for testing genetic anticipation: looking for an answer in Lynch syndrome. Genetic epidemiology, 34(7), 756-68. URL http://www.ncbi.nlm.nih.gov/pubmed/20878717.

Bowers, J., Sokolsky, T., Quach, T., \& Alani, E. (1999). A mutation in the MSH6 subunit of the Saccharomyces cerevisiae MSH2-MSH6 complex disrupts mismatch recognition. The Journal of biological chemistry, 274(23), 16115-25. URL http: //www.ncbi.nlm.nih.gov/pubmed/10347163.

Boyer, J. C. \& Farber, R. A. (1998). Mutation rate of a microsatellite sequence in normal human fibroblasts. Cancer research, 58(17), 3946-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/9731507.

Boyer, J. C., Umar, A., Risinger, J. I., Lipford, J. R., Kane, M., Yin, S., Barrett, J. C., Kolodner, R. D., \& Kunkel, T. A. (1995). Microsatellite instability, mismatch repair deficiency, and genetic defects in human cancer cell lines. Cancer research, 55(24), 6063-70. URL http: //www.ncbi.nlm.nih.gov/pubmed/8521394.

Bozzao, C., Lastella, P., Ponz de Leon, M., Pedroni, M., Di Gregorio, C., D’Ovidio, F. D., Resta, N., Prete, F., Guanti, G., \& Stella, A. (2011a). Analysis of telomere dynamics in peripheral blood cells from patients with Lynch syndrome. Cancer, 117(18), 4325-35. URL http://www.ncbi.nlm.nih.gov/pubmed/21387278.

Bozzao, C., Lastella, P., \& Stella, A. (2011b). Anticipation in lynch syndrome: where we are where we go. Current genomics, 12(7), 451-65. URL http: //www.ncbi.nlm.nih.gov/pubmed/22547953.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry, 72, 248-54. URL http://www.ncbi.nlm.nih.gov/pubmed/942051.

Bridge, G., Rashid, S., \& Martin, S. a. (2014). DNA mismatch repair and oxidative DNA damage: implications for cancer biology and treatment. Cancers, 6(3), 1597-614. URL http://www.ncbi.nlm.nih.gov/pulmed/25099886.

Britt-Compton, B., Capper, R., Rowson, J., \& Baird, D. M. (2009a). Short telomeres are preferentially elongated by telomerase in human cells. FEBS letters, 583(18), 3076-80. URL http://www.ncbi.nlm.nih.gov/pubmed/19716824.

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(5), 725-33. URL http://www.ncbi.nlm.nih.gov/pubmed/16421168.

Britt-Compton, B., Wyllie, F., Rowson, J., Capper, R., E Jones, R., \& M Baird, D. (2009b). Telomere dynamics during replicative senescence are not directly modulated by conditions of oxidative stress in IMR90 fibroblast cells. Biogerontology, 10(6), 683-93. URL http: //www.ncbi.nlm.nih.gov/pubmed/19214769.

Broer, L., Codd, V., Nyholt, D. R., Deelen, J., Mangino, M., Willemsen, G., Albrecht, E., Amin, N., Beekman, M., de Geus, E. J. C., Henders, A., Nelson, C. P., Steves, C. J., Wright, M. J., de Craen, A. J. M., Isaacs, A., Matthews, M., Moayyeri, A., Montgomery, G. W., Oostra, B. a., Vink, J. M., Spector, T. D., Slagboom, P. E., Martin, N. G., Samani, N. J., van Duijn, C. M., \& Boomsma, D. I. (2013). Meta-analysis of telomere length in 19,713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect. European journal of human genetics : EJHG, 21(10), 1163-8. URL http://www.ncbi.nlm.nih.gov/pubmed/23321625.

Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., \& Lindblom, A. (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature, 368(6468), 258-61. URL http://www.ncbi.nlm.nih.gov/pubmed/8145827.

Brümmendorf, T. H., Maciejewski, J. P., Mak, J., Young, N. S., \& Lansdorp, P. M. (2001). Telomere length in leukocyte subpopulations of patients with aplastic anemia. Blood, 97(4), 895-900. URL http://www.ncbi.nlm.nih.gov/pubmed/11159514.

Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A., \& Reddel, R. R. (1997). Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. Nature medicine, 3(11), 1271-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/9359704.

Buecher, B., Cacheux, W., Rouleau, E., Dieumegard, B., Mitry, E., \& Lièvre, A. (2012). Role of microsatellite instability in the management of colorectal cancers. Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver. URL http://www.ncbi.nlm.nih.gov/pubmed/23195666.

Campbell, I. M., Shaw, C. A., Stankiewicz, P., \& Lupski, J. R. (2015). Somatic mosaicism: implications for disease and transmission genetics. Trends in genetics : TIG, 31(7), 38292. URL http://www.ncbi.nlm.nih.gov/pubmed/25910407.

Campbell, M. R., Wang, Y., Andrew, S. E., \& Liu, Y. (2006). Msh2 deficiency leads to chromosomal abnormalities, centrosome amplification, and telomere capping defect. Oncogene, 25(17), 2531-6. URL http://www.ncbi.nlm.nih.gov/pubmed/16331258.

Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature, 487(7407), 330-7. URL http: //www.ncbi.nlm.nih.gov/pubmed/22810696.

Cawthon, R. M. (2002). Telomere measurement by quantitative PCR. Nucleic acids research, 30(10), e47. URL http://www.ncbi.nlm.nih.gov/pubmed/19129229.

Cawthon, R. M. (2009). Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic acids research, 37(3), e21. URL http: //www.ncbi.nlm.nih.gov/pubmed/19129229.

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 (London, England), 361(9355), 393-5. URL http: //www.ncbi.nlm.nih.gov/pubmed/12573379.

Centelles, J. J. (2012). General aspects of colorectal cancer. ISRN oncology, 2012, 139268. URL http://www.ncbi.nlm.nih.gov/pubmed/23209942.

Chadeneau, C., Hay, K., Hirte, H. W., Gallinger, S., \& Bacchetti, S. (1995). Telomerase activity associated with acquisition of malignancy in human colorectal cancer. Cancer research, 55(12), 2533-6. URL http://www.ncbi.nlm.nih.gov/pubmed/7780964.

Chang, S.-C., Lin, J.-K., Lin, T.-C., \& Liang, W.-Y. (2005). Loss of heterozygosity: an independent prognostic factor of colorectal cancer. World journal of gastroenterology : WJG, 11(6), 778-84. URL http://www.ncbi.nlm.nih.gov/pubmed/15682467.

Chen, J.-M., Cooper, D. N., Chuzhanova, N., Férec, C., \& Patrinos, G. P. (2007). Gene conversion: mechanisms, evolution and human disease. Nature reviews. Genetics, 8(10), 762-75. URL http://www.ncbi.nlm.nih.gov/pubmed/17846636.

Chen, W., Kimura, M., Kim, S., Cao, X., Srinivasan, S. R., Berenson, G. S., Kark, J. D., \& Aviv, A. (2011). Longitudinal versus cross-sectional evaluations of leukocyte telomere length dynamics: age-dependent telomere shortening is the rule. The journals of gerontology. Series A, Biological sciences and medical sciences, 66(3), 312-9. URL http://www.ncbi.nlm.nih.gov/pubmed/21310811.

Chen, X., Velez, J. C., Barbosa, C., Pepper, M., Andrade, A., Stoner, L., De Vivo, I., Gelaye, B., \& Williams, M. A. (2015). Smoking and perceived stress in relation to short salivary telomere length among caregivers of children with disabilities. Stress (Amsterdam, Netherlands), 18(1), 20-8. URL http://www.ncbi.nlm.nih.gov/pubmed/25256607.

Cheng, a. J., Tang, R., Wang, J. Y., See, L. C., \& Wang, T. C. (1998). Possible role of telomerase activation in the cancer predisposition of patients with hereditary nonpolyposis colorectal cancers. Journal of the National Cancer Institute, 90(4), 316-21. URL http://www.ncbi.nlm.nih.gov/pubmed/9486818.

Cheng, G., Kong, F., Luan, Y., Sun, C., Wang, J., Zhang, L., Jiang, B., Qi, T., Zhao, J., Zheng, C., \& Xu, D. (2013). Differential shortening rate of telomere length in the development of human fetus. Biochemical and biophysical research communications, 442(1-2), 112-5. URL http://www.ncbi.nlm.nih.gov/pubmed/24246679.

Cherkas, L. F., Aviv, A., Valdes, A. M., Hunkin, J. L., Gardner, J. P., Surdulescu, G. L., Kimura, M., \& Spector, T. D. (2006). The effects of social status on biological aging as measured by white-blood-cell telomere length. Aging cell, 5(5), 361-5. URL http: //www.ncbi.nlm.nih.gov/pubmed/16856882.

Cherkas, L. F., Hunkin, J. L., Kato, B. S., Richards, J. B., Gardner, J. P., Surdulescu, G. L., Kimura, M., Lu, X., Spector, T. D., \& Aviv, A. (2008). The association between physical activity in leisure time and leukocyte telomere length. Archives of internal medicine, 168(2), 154-8. URL http://www.ncbi.nlm.nih.gov/pubmed/18227361.

Chiang, Y. J., Calado, R. T., Hathcock, K. S., Lansdorp, P. M., Young, N. S., \& Hodes, R. J. (2010). Telomere length is inherited with resetting of the telomere set-point. Proceedings of the National Academy of Sciences of the United States of America, 107(22), 10148-53. URL http://www.ncbi.nlm.nih.gov/pubmed/20479226.

Chiappin, S., Antonelli, G., Gatti, R., \& De Palo, E. F. (2007). Saliva specimen: a new laboratory tool for diagnostic and basic investigation. Clinica chimica acta; international journal of clinical chemistry, 383(1-2), 30-40. URL http: //www.ncbi.nlm.nih.gov/pubmed/17512510.

Chiodi, I. \& Mondello, C. (2012). Telomere-independent functions of telomerase in nuclei, cytoplasm, and mitochondria. Frontiers in oncology, 2(September), 133. URL http: //www.ncbi.nlm.nih.gov/pubmed/23061047.

Chow, T. T., Zhao, Y., Mak, S. S., Shay, J. W., \& Wright, W. E. (2012). Early and late steps in telomere overhang processing in normal human cells: the position of the final RNA primer drives telomere shortening. Genes \& development, 26(11), 1167-78. URL http: //www.ncbi.nlm.nih.gov/pubmed/22661228.

Cilli, P., Minoprio, A., Bossa, C., Bignami, M., \& Mazzei, F. (2015). Formation and Repair of Mismatches Containing Ribonucleotides and Oxidized Bases at Repeated DNA Sequences. The Journal of biological chemistry, 290(43), 26259-69. URL http: //www.ncbi.nlm.nih.gov/pubmed/26338705.

Cimprich, K. a. \& Cortez, D. (2008). ATR: an essential regulator of genome integrity. Nature reviews. Molecular cell biology, 9(8), 616-27. URL http: //www.ncbi.nlm.nih.gov/pubmed/18594563.

Clark, N., Wu, X., \& Her, C. (2013). MutS Homologues hMSH4 and hMSH5: Genetic Variations, Functions, and Implications in Human Diseases. Current genomics, 14(2), 81-90. URL http://www.ncbi.nlm.nih.gov/pubmed/24082819.

Clatworthy, J. P. \& Subramanian, V. (2001). Stem cells and the regulation of proliferation, differentiation and patterning in the intestinal epithelium: emerging insights from gene expression patterns, transgenic and gene ablation studies. Mechanisms of development, 101(1-2), 3-9. URL http://www.ncbi.nlm.nih.gov/pubmed/11231054.

Claverys, J. P. \& Lacks, S. a. (1986). Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. Microbiological reviews, 50(2), 133-65. URL http: //www.ncbi.nlm.nih.gov/pubmed/3523187.

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 16 p and 16q. Human molecular genetics, 8(9), 1637-46. URL http: //www.ncbi.nlm.nih.gov/pubmed/10441326.

Coluzzi, E., Colamartino, M., Cozzi, R., Leone, S., Meneghini, C., O’Callaghan, N., \& Sgura, A. (2014). Oxidative stress induces persistent telomeric DNA damage responsible for nuclear morphology change in mammalian cells. PloS one, 9(10), e110963. URL http://www.ncbi.nlm.nih.gov/pubmed/25354277.

Conde-Pérezprina, J. C., León-Galván, M. Á., \& Konigsberg, M. (2012). DNA mismatch repair system: repercussions in cellular homeostasis and relationship with aging. Oxidative medicine and cellular longevity, 2012, 728430. URL http: //www.ncbi.nlm.nih.gov/pubmed/23213348.

Cong, Y.-s., Wright, W. E., \& Shay, J. W. (2002). Human telomerase and its regulation. Microbiology and molecular biology reviews : MMBR, 66(3), 407-25, table of contents. URL http://www.ncbi.nlm.nih.gov/pubmed/12208997.

Conomos, D., Stutz, M. D., Hills, M., Neumann, A. A., Bryan, T. M., Reddel, R. R., \& Pickett, H. A. (2012). Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. The Journal of cell biology, 199(6), 893-906. URL http://www.ncbi.nlm.nih.gov/pubmed/23229897http: //www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3518223.

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(1), 332-42. URL http: //www.ncbi.nlm.nih.gov/pubmed/11739745.

Coolbaugh-Murphy, M., Maleki, A., Ramagli, L., Frazier, M., Lichtiger, B., Monckton, D. G., Siciliano, M. J., \& Brown, B. W. (2004). Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR. Genomics, 84(2), 419-30. URL http: //www.ncbi.nlm.nih.gov/pubmed/15234004.

Coolbaugh-Murphy, M. I., Xu, J.-P., Ramagli, L. S., Ramagli, B. C., Brown, B. W., Lynch, P. M., Hamilton, S. R., Frazier, M. L., \& Siciliano, M. J. (2010). Microsatellite instability in the peripheral blood leukocytes of HNPCC patients. Human mutation, 31(3), 317-24. URL http://www.ncbi.nlm.nih.gov/pubmed/20052760.

Cooper, D. L., Lahue, R. S., \& Modrich, P. (1993). Methyl-directed mismatch repair is bidirectional. The Journal of biological chemistry, 268(16), 11823-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/8389365.

Crouse, G. F. (2016). Non-canonical actions of mismatch repair. DNA repair, 38, 102-9. URL http://www.ncbi.nlm.nih.gov/pubmed/26698648.

Cunningham, J. M., Johnson, R. a., Litzelman, K., Skinner, H. G., Seo, S., Engelman, C. D., Vanderboom, R. J., Kimmel, G. W., Gangnon, R. E., Riegert-Johnson, D. L., Baron, J. a., Potter, J. D., Haile, R., Buchanan, D. D., Jenkins, M. a., Rider, D. N., Thibodeau, S. N., Petersen, G. M., \& Boardman, L. a. (2013). Telomere length varies by DNA extraction method: implications for epidemiologic research. Cancer epidemiology, biomarkers \& prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 22(11), 2047-54. URL http: //www.ncbi.nlm.nih.gov/pubmed/24019396.

Cusanelli, E., Romero, C. A. P., \& Chartrand, P. (2013). Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. Molecular cell, 51(6), 780-91. URL http: //www.ncbi.nlm.nih.gov/pubmed/24074956.

Damm, K., Hemmann, U., Garin-Chesa, P., Hauel, N., Kauffmann, I., Priepke, H., Niestroj, C., Daiber, C., Enenkel, B., Guilliard, B., Lauritsch, I., Müller, E., Pascolo, E., Sauter, G., Pantic, M., Martens, U. M., Wenz, C., Lingner, J., Kraut, N., Rettig, W. J., \& Schnapp, A. (2001). A highly selective telomerase inhibitor limiting human cancer cell proliferation. The EMBO journal, 20(24), 6958-68. URL http: //www.ncbi.nlm.nih.gov/pubmed/11742973.

Daniali, L., Benetos, A., Susser, E., Kark, J. D., Labat, C., Kimura, M., Desai, K., Granick, M., \& Aviv, A. (2013). Telomeres shorten at equivalent rates in somatic tissues of adults. Nature communications, 4, 1597. URL http: //www.ncbi.nlm.nih.gov/pubmed/23511462.

Dao, V. \& Modrich, P. (1998). Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. The Journal of biological chemistry, 273(15), 9202-7. URL http://www.ncbi.nlm.nih.gov/pubmed/9535911.
de Lange, T. (2002). Protection of mammalian telomeres. Oncogene, 21(4), 532-40. URL http://www.ncbi.nlm.nih.gov/pubmed/11850778.
de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. Genes \& development, 19(18), 2100-10. URL http: //www.ncbi.nlm.nih.gov/pubmed/16166375.
de Lange, T. (2009). How telomeres solve the end-protection problem. Science (New York, N.Y.), 326(5955), 948-52. URL http: / /www.ncbi.nlm.nih.gov/pubmed/19965504.

De Meyer, T., Rietzschel, E. R., De Buyzere, M. L., De Bacquer, D., Van Criekinge, W., De Backer, G. G., Gillebert, T. C., Van Oostveldt, P., Bekaert, S., \& investigators, A. (2007). Paternal age at birth is an important determinant of offspring telomere length. Human molecular genetics, 16(24), 3097-102. URL http: //www.ncbi.nlm.nih.gov/pubmed/17881651.
de Wind, N., Dekker, M., Claij, N., Jansen, L., van Klink, Y., Radman, M., Riggins, G., van der Valk, M., van't Wout, K., \& te Riele, H. (1999). HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. Nature genetics, 23(3), 359-62. URL http: / /www.ncbi.nlm.nih.gov/pubmed/10545954.

Debacq-Chainiaux, F., Erusalimsky, J. D., Campisi, J., \& Toussaint, O. (2009). Protocols to detect senescence-associated beta-galactosidase (SA-betagal) activity, a biomarker of senescent cells in culture and in vivo. Nature protocols, 4(12), 1798-806. URL http: //www.ncbi.nlm.nih.gov/pubmed/20010931.

Deferme, L., Briedé, J. J., Claessen, S. M. H., Cavill, R., \& Kleinjans, J. C. S. (2015). Cell line-specific oxidative stress in cellular toxicity: A toxicogenomics-based comparison between liver and colon cell models. Toxicology in vitro : an international journal published in association with BIBRA, 29(5), 845-55. URL http: //www.ncbi.nlm.nih.gov/pubmed/25800948.

Degasperi, A., Birtwistle, M. R., Volinsky, N., Rauch, J., Kolch, W., \& Kholodenko, B. N. (2014). Evaluating strategies to normalise biological replicates of Western blot data. PloS one, 9(1), e87293. URL http://www.ncbi.nlm.nih.gov/pubmed/24475266.

Delhommeau, F., Thierry, A., Feneux, D., Lauret, E., Leclercq, E., Courtier, M. H., Sainteny, F., Vainchenker, W., \& Bennaceur-Griscelli, A. (2002). Telomere dysfunction and telomerase reactivation in human leukemia cell lines after telomerase inhibition by the expression of a dominant-negative hTERT mutant. Oncogene, 21(54), 8262-71. URL http://www.ncbi.nlm.nih.gov/pubmed/12447689.

Deschoolmeester, V., Baay, M., Specenier, P., Lardon, F., \& Vermorken, J. B. (2010). A review of the most promising biomarkers in colorectal cancer: one step closer to targeted therapy. The oncologist, 15(7), 699-731. URL http: //www.ncbi.nlm.nih.gov/pubmed/20584808.

DiFrancesco, R., Bhatnagar, S. K., Brown, A., \& Bessman, M. J. (1984). The interaction of DNA polymerase III and the product of the Escherichia coli mutator gene, mutD. The Journal of biological chemistry, 259(9), 5567-73. URL http: //www.ncbi.nlm.nih.gov/pubmed/6325441.

Dizdaroglu, M. (2012). Oxidatively induced DNA damage: mechanisms, repair and disease. Cancer letters, 327(1-2), 26-47. URL http: //www.ncbi.nlm.nih.gov/pubmed/22293091.

Drost, M., Lützen, A., van Hees, S., Ferreira, D., Calléja, F., Zonneveld, J. B. M., Nielsen, F. C., Rasmussen, L. J., \& de Wind, N. (2013). Genetic screens to identify pathogenic gene variants in the common cancer predisposition Lynch syndrome. Proceedings of the National Academy of Sciences of the United States of America, 110(23), 9403-8. URL http://www.ncbi.nlm.nih.gov/pubmed/23690608.

Drummond, J. T., Li, G. M., Longley, M. J., \& Modrich, P. (1995). Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. Science (New York, N.Y.), 268(5219), 1909-12. URL http://www.ncbi.nlm.nih.gov/pubmed/7604264.

Duggan, C., Risques, R., Alfano, C., Prunkard, D., Imayama, I., Holte, S., Baumgartner, K., Baumgartner, R., Bernstein, L., Ballard-Barbash, R., Rabinovitch, P., \& McTiernan, A. (2014). Change in peripheral blood leukocyte telomere length and mortality in breast cancer survivors. Journal of the National Cancer Institute, 106(4), dju035. URL http: //www.ncbi.nlm.nih.gov/pubmed/24627273.

Dumont, P., Royer, V., Pascal, T., Dierick, J. F., Chainiaux, F., Frippiat, C., de Magalhaes, J. P., Eliaers, F., Remacle, J., \& Toussaint, O. (2001). Growth kinetics rather than stress accelerate telomere shortening in cultures of human diploid fibroblasts in oxidative stress-induced premature senescence. FEBS letters, 502(3), 109-12. URL http://www.ncbi.nlm.nih.gov/pubmed/11583109.

Ehrlenbach, S., Willeit, P., Kiechl, S., Willeit, J., Reindl, M., Schanda, K., Kronenberg, F., \& Brandstätter, A. (2009). Influences on the reduction of relative telomere length over 10 years in the population-based Bruneck Study: introduction of a well-controlled highthroughput assay. International journal of epidemiology, 38(6), 1725-34. URL http: //www.ncbi.nlm.nih.gov/pubmed/19666704.

Eisenberg, D. T. a. (2014). Inconsistent inheritance of telomere length (TL): is offspring TL more strongly correlated with maternal or paternal TL? European journal of human genetics : EJHG, 22(1), 8-9. URL http://www.ncbi.nlm.nih.gov/pubmed/24022299.

Eisenberg, D. T. a., Hayes, M. G., \& Kuzawa, C. W. (2012). Delayed paternal age of reproduction in humans is associated with longer telomeres across two generations of descendants. Proceedings of the National Academy of Sciences of the United States of America, 109(26), 10251-6. URL http://www.ncbi.nlm.nih.gov/pubmed/22689985.

El-Daly, H., Kull, M., Zimmermann, S., Pantic, M., Waller, C. F., \& Martens, U. M. (2005). Selective cytotoxicity and telomere damage in leukemia cells using the telomerase inhibitor BIBR1532. Blood, 105(4), 1742-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/15507522.

El Daly, H. \& Martens, U. M. (2007). Telomerase inhibition and telomere targeting in hematopoietic cancer cell lines with small non-nucleosidic synthetic compounds (BIBR1532). Methods in molecular biology (Clifton, N.J.), 405(1), 47-60. URL http: //www.ncbi.nlm.nih.gov/pubmed/18369817.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nature reviews. Genetics, 5(6), 435-45. URL http: //www.ncbi.nlm.nih.gov/pubmed/15153996.

Elsayed, F. A., Kets, C. M., Ruano, D., van den Akker, B., Mensenkamp, A. R., Schrumpf, M., Nielsen, M., Wijnen, J. T., Tops, C. M., Ligtenberg, M. J., Vasen, H. F. A., Hes, F. J., Morreau, H., \& van Wezel, T. (2015). Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. European journal of human genetics : EJHG, 23(8), 1080-4. URL http://www.ncbi.nlm.nih.gov/pubmed/25370038.

Epel, E. S., Merkin, S. S., Cawthon, R., Blackburn, E. H., Adler, N. E., Pletcher, M. J., \& Seeman, T. E. (2009). The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. Aging, 1(1), 81-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/20195384.

Fan, J., Cai, H., Yang, S., Yan, L., \& Tan, W. (2008). Comparison between the effects of normoxia and hypoxia on antioxidant enzymes and glutathione redox state in ex vivo culture of CD34(+) cells. Comparative biochemistry and physiology. Part B, Biochemistry \& molecular biology, 151(2), 153-8. URL http://www.ncbi.nlm.nih.gov/pubmed/18625331.

Farzaneh-Far, R., Lin, J., Epel, E., Lapham, K., Blackburn, E., \& Whooley, M. A. (2010). Telomere length trajectory and its determinants in persons with coronary artery disease: longitudinal findings from the heart and soul study. PloS one, 5(1), e8612. URL http: //www.ncbi.nlm.nih.gov/pubmed/20072607.

Fearon, E. R. \& Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell, 61(5), 759-67. URL http: / /www.ncbi.nlm.nih.gov/pubmed/2188735.

Ferlin, a., Rampazzo, E., Rocca, M. S., Keppel, S., Frigo, a. C., De Rossi, a., \& Foresta, C. (2013). In young men sperm telomere length is related to sperm number and parental age. Human reproduction (Oxford, England), 28(12), 3370-6. URL http: //www.ncbi.nlm.nih.gov/pubmed/24166593.

Fiorentini, P., Huang, K. N., Tishkoff, D. X., Kolodner, R. D., \& Symington, L. S. (1997). Exonuclease I of Saccharomyces cerevisiae functions in mitotic recombination in vivo and in vitro. Molecular and cellular biology, 17(5), 2764-73. URL http: //www.ncbi.nlm.nih.gov/pubmed/9111347.

Fishel, R. (1998). Mismatch repair, molecular switches, and signal transduction. Genes \& development, 12(14), 2096-101. URL http: //www.ncbi.nlm.nih.gov/pubmed/9679053.

Fishel, R. \& Kolodner, R. D. (1995). Identification of mismatch repair genes and their role in the development of cancer. Current opinion in genetics \& development, 5(3), 382-95. URL http://www.ncbi.nlm.nih.gov/pubmed/7549435.

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. a., Garber, J., Kane, M., \& Kolodner, R. (1994). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell, 77(1), 1 p following 166. URL http://www.ncbi.nlm.nih.gov/pubmed/8156592.

Fitzpatrick, A. L., Kronmal, R. a., Kimura, M., Gardner, J. P., Psaty, B. M., Jenny, N. S., Tracy, R. P., Hardikar, S., \& Aviv, A. (2011). Leukocyte telomere length and mortality in the Cardiovascular Health Study. The journals of gerontology. Series A, Biological sciences and medical sciences, 66(4), 421-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/21289018.

Flanagan, L., Meyer, M., Fay, J., Curry, S., Bacon, O., Duessmann, H., John, K., Boland, K. C., McNamara, D. A., Kay, E. W., Bantel, H., Schulze-Bergkamen, H., \& Prehn, J. H. M. (2016). Low levels of Caspase-3 predict favourable response to 5FU-based chemotherapy in advanced colorectal cancer: Caspase-3 inhibition as a therapeutic approach. Cell death \& disease, 7(2), e2087. URL http://www.ncbi.nlm.nih.gov/pubmed/26844701.

Fleisig, H. B., Hukezalie, K. R., Thompson, C. a. H., Au-Yeung, T. T. T., Ludlow, a. T., Zhao, C. R., \& Wong, J. M. Y. (2016). Telomerase reverse transcriptase expression protects transformed human cells against DNA-damaging agents, and increases tolerance to chromosomal instability. Oncogene, 35(2), 218-27. URL http: //www.ncbi.nlm.nih.gov/pubmed/25893297.

Flood, C. L., Rodriguez, G. P., Bao, G., Shockley, A. H., Kow, Y. W., \& Crouse, G. F. (2015). Replicative DNA polymerase $\delta$ but not $\epsilon$ proofreads errors in Cis and in Trans. PLoS genetics, 11(3), e1005049. URL http://www.ncbi.nlm.nih.gov/pubmed/25742645.

Fodde, R., Kuipers, J., Rosenberg, C., Smits, R., Kielman, M., Gaspar, C., van Es, J. H., Breukel, C., Wiegant, J., Giles, R. H., \& Clevers, H. (2001a). Mutations in the APC tumour suppressor gene cause chromosomal instability. Nature cell biology, 3(4), 433-8. URL http://www.ncbi.nlm.nih.gov/pubmed/11283620.

Fodde, R., Smits, R., \& Clevers, H. (2001b). APC, signal transduction and genetic instability in colorectal cancer. Nature reviews. Cancer, 1(1), 55-67. URL http: //www.ncbi.nlm.nih.gov/pubmed/11900252.

Frenck, R. W., Blackburn, E. H., \& Shannon, K. M. (1998). The rate of telomere sequence loss in human leukocytes varies with age. Proceedings of the National Academy of Sciences of the United States of America, 95(10), 5607-10. URL http: //www.ncbi.nlm.nih.gov/pubmed/9576930.

Frescas, D. \& de Lange, T. (2014). A TIN2 dyskeratosis congenita mutation causes telomerase-independent telomere shortening in mice. Genes \& development, 28(2), 15366. URL http://www.ncbi.nlm.nih.gov/pubmed/24449270.

Friedrich, U., Griese, E., Schwab, M., Fritz, P., Thon, K., \& Klotz, U. (2000). Telomere length in different tissues of elderly patients. Mechanisms of ageing and development, 119(3), 89-99. URL http://www.ncbi.nlm.nih.gov/pubmed/11080530.

Fujii, H. \& Shimada, T. (1989). Isolation and characterization of cDNA clones derived from the divergently transcribed gene in the region upstream from the human dihydrofolate reductase gene. The Journal of biological chemistry, 264(17), 10057-64. URL http://www.ncbi.nlm.nih.gov/pubmed/2722860.

Fujisawa, M., Tanaka, H., Tatsumi, N., Okada, H., Arakawa, S., \& Kamidono, S. (1998). Telomerase activity in the testis of infertile patients with selected causes. Human reproduction (Oxford, England), 13(6), 1476-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/9688375.

Fujiwara, T., Stolker, J. M., Watanabe, T., Rashid, A., Longo, P., Eshleman, J. R., Booker, S., Lynch, H. T., Jass, J. R., Green, J. S., Kim, H., Jen, J., Vogelstein, B., \& Hamilton, S. R. (1998). Accumulated clonal genetic alterations in familial and sporadic colorectal carcinomas with widespread instability in microsatellite sequences. The American journal of pathology, 153(4), 1063-78. URL http://www.ncbi.nlm.nih.gov/pubmed/9777938.

Fumagalli, M., Rossiello, F., Mondello, C., \& d’Adda di Fagagna, F. (2014). Stable cellular senescence is associated with persistent DDR activation. PloS one, 9(10), e110969. URL http://www.ncbi.nlm.nih.gov/pubmed/25340529.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., \& Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics (Oxford, England), 28(20), 2678-9. URL http://www.ncbi.nlm.nih.gov/pubmed/22914218.

Gardner, J. P., Li, S., Srinivasan, S. R., Chen, W., Kimura, M., Lu, X., Berenson, G. S., \& Aviv, A. (2005). Rise in insulin resistance is associated with escalated telomere attrition. Circulation, 111(17), 2171-7. URL http://www.ncbi.nlm.nih.gov/pubmed/15851602.

Gardner, M., Bann, D., Wiley, L., Cooper, R., Hardy, R., Nitsch, D., Martin-Ruiz, C., Shiels, P., Sayer, A. A., Barbieri, M., Bekaert, S., Bischoff, C., Brooks-Wilson, A., Chen, W., Cooper, C., Christensen, K., De Meyer, T., Deary, I., Der, G., Diez Roux, A., Fitzpatrick, A., Hajat, A., Halaschek-Wiener, J., Harris, S., Hunt, S. C., Jagger, C., Jeon, H.-S., Kaplan, R., Kimura, M., Lansdorp, P., Li, C., Maeda, T., Mangino, M., Nawrot, T. S., Nilsson, P., Nordfjall, K., Paolisso, G., Ren, F., Riabowol, K., Robertson, T., Roos, G., Staessen, J. A., Spector, T., Tang, N., Unryn, B., van der Harst, P., Woo, J., Xing, C., Yadegarfar, M. E., Park, J. Y., Young, N., Kuh, D., von Zglinicki, T., Ben-Shlomo, Y., \& study team, H. (2014). Gender and telomere length: systematic review and meta-analysis. Experimental gerontology, 51(1), 15-27. URL http://www.ncbi.nlm.nih.gov/pubmed/24365661.

Garre, P., Martín, L., Bando, I., Tosar, A., Llovet, P., Sanz, J., Romero, A., de la Hoya, M., Díaz-Rubio, E., \& Caldés, T. (2014). Cancer risk and overall survival in mismatch repair proficient hereditary non-polyposis colorectal cancer, Lynch syndrome and sporadic colorectal cancer. Familial cancer, 13(1), 109-19. URL http: //www.ncbi.nlm.nih.gov/pubmed/24061861.

Garrido-Ramos, M. (Ed.) (2012). Repetitive DNA, vol. 7 of Genome Dynamics. Granada: S. Karger AG. URL http://www.karger.com/Book/Home/256867.

Gary, R. K. \& Kindell, S. M. (2005). Quantitative assay of senescence-associated betagalactosidase activity in mammalian cell extracts. Analytical biochemistry, 343(2), 329-34. URL http://www.ncbi.nlm.nih.gov/pubmed/16004951.

Geiersbach, K. B. \& Samowitz, W. S. (2011). Microsatellite instability and colorectal cancer. Archives of pathology \& laboratory medicine, 135(10), 1269-77. URL http: //www.ncbi.nlm.nih.gov/pubmed/21970482.

Genschel, J., Bazemore, L. R., \& Modrich, P. (2002). Human exonuclease I is required for 5' and 3' mismatch repair. The Journal of biological chemistry, 277(15), 13302-11. URL http://www.ncbi.nlm.nih.gov/pubmed/11809771.

Ghodgaonkar, M. M., Lazzaro, F., Olivera-Pimentel, M., Artola-Borán, M., Cejka, P., Reijns, M. a., Jackson, A. P., Plevani, P., Muzi-Falconi, M., \& Jiricny, J. (2013). Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. Molecular cell, 50(3), 323-32. URL http: //www.ncbi.nlm.nih.gov/pubmed/23603115.

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., \& McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome research, 25(11), 1750-6. URL http: //www.ncbi.nlm.nih.gov/pubmed/26447147.

Graakjaer, J., Der-Sarkissian, H., Schmitz, A., Bayer, J., Thomas, G., Kolvraa, S., \& LondoñoVallejo, J.-A. (2006). Allele-specific relative telomere lengths are inherited. Human genetics, 119(3), 344-50. URL http://www.ncbi.nlm.nih.gov/pubmed/16440201.

Graakjaer, J., Pascoe, L., Der-Sarkissian, H., Thomas, G., Kolvraa, S., Christensen, K., \& Londoño-Vallejo, J.-A. (2004). The relative lengths of individual telomeres are defined in the zygote and strictly maintained during life. Aging cell, 3(3), 97-102. URL http: //www.ncbi.nlm.nih.gov/pubmed/15153177.

Gradia, S., Acharya, S., \& Fishel, R. (1997). The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch. Cell, 91(7), 995-1005. URL http://www.ncbi.nlm.nih.gov/pubmed/9428522.

Gradia, S., Acharya, S., \& Fishel, R. (2000). The role of mismatched nucleotides in activating the hMSH2-hMSH6 molecular switch. The Journal of biological chemistry, 275(6), 392230. URL http://www.ncbi.nlm.nih.gov/pubmed/10660545.

Greider, C. W. (1996). Telomere length regulation. Annual review of biochemistry, 65, 337-65. URL http://www.ncbi.nlm.nih.gov/pubmed/8811183.

Greider, C. W. \& Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell, $43(2 \mathrm{Pt} 1), 405-13$. URL http: //www.ncbi.nlm.nih.gov/pubmed/3907856.

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, a., Moss, H., \& de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. Cell, 97(4), 503-14. URL http://www.ncbi.nlm.nih.gov/pubmed/10338214.

Gu, B., Bessler, M., \& Mason, P. J. (2009). Dyskerin, telomerase and the DNA damage response. Cell cycle (Georgetown, Tex.), 8(1), 6-10. URL http: //www.ncbi.nlm.nih.gov/pubmed/19106610.

Haber, L. T. \& Walker, G. C. (1991). Altering the conserved nucleotide binding motif in the Salmonella typhimurium MutS mismatch repair protein affects both its ATPase and mismatch binding activities. The EMBO journal, 10(9), 2707-15. URL http: //www.ncbi.nlm.nih.gov/pu.bmed/1651234.

Hahn, W. C., Stewart, S. a., Brooks, M. W., York, S. G., Eaton, E., Kurachi, a., Beijersbergen, R. L., Knoll, J. H., Meyerson, M., \& Weinberg, R. a. (1999). Inhibition of telomerase limits the growth of human cancer cells. Nature medicine, 5(10), 1164-70. URL http://www.ncbi.nlm.nih.gov/pubmed/10502820.

Halaschek-Wiener, J., Vulto, I., Fornika, D., Collins, J., Connors, J. M., Le, N. D., Lansdorp, P. M., \& Brooks-Wilson, A. (2008). Reduced telomere length variation in healthy oldest old. Mechanisms of ageing and development, 129(11), 638-41. URL http: //www.ncbi.nlm.nih.gov/pubmed/18765247.

Hanford, M. G., Rushton, B. C., Gowen, L. C., \& Farber, R. a. (1998). Microsatellite mutation rates in cancer cell lines deficient or proficient in mismatch repair. Oncogene, 16(18), 238993. URL http://www.ncbi.nlm.nih.gov/pubmed/9620556.

Harfe, B. D. \& Jinks-Robertson, S. (2000a). DNA mismatch repair and genetic instability. Annual review of genetics, 34, 359-399. URL http: //www.ncbi.nlm.nih.gov/pubmed/11092832.

Harfe, B. D. \& Jinks-Robertson, S. (2000b). Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in Saccharomyces cerevisiae. Genetics, 156(2), 571-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/11014807.

Harley, C. B., Futcher, A. B., \& Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature, 345(6274), 458-60. URL http: //www.ncbi.nlm.nih.gov/pubmed/2342578.

Hassen, S., Ali, A. A., Kilaparty, S. P., Al-Anbaky, Q. A., Majeed, W., Boman, B. M., Fields, J. Z., \& Ali, N. (2016). Interdependence of DNA mismatch repair proteins MLH1 and MSH2 in apoptosis in human colorectal carcinoma cell lines. Molecular and cellular biochemistry, 412(1-2), 297-305. URL http://www.ncbi.nlm.nih.gov/pubmed/26728996.

Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D. K., \& Allshire, R. C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. Nature, 346(6287), 866-8. URL http://www.ncbi.nlm.nih.gov/pubmed/2392154.

Hayashi, M. T., Cesare, A. J., Rivera, T., \& Karlseder, J. (2015). Cell death during crisis is mediated by mitotic telomere deprotection. Nature, 522(7557), 492-6. URL http: //www.ncbi.nlm.nih.gov/pubmed/26108857.

Haye, J. E. \& Gammie, A. E. (2015). The Eukaryotic Mismatch Recognition Complexes Track with the Replisome during DNA Synthesis. PLoS genetics, 11(12), e1005719. URL http: //www.ncbi.nlm.nih.gov/pubmed/26684201.

Hayflick, L. \& Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. Experimental cell research, 25, 585-621. URL http: //www.ncbi.nlm.nih.gov/pubmed/13905658.

Hegan, D. C., Narayanan, L., Jirik, F. R., Edelmann, W., Liskay, R. M., \& Glazer, P. M. (2006). Differing patterns of genetic instability in mice deficient in the mismatch repair genes Pms2, Mlh1, Msh2, Msh3 and Msh6. Carcinogenesis, 27(12), 2402-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/16728433.

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(1), 67-77. URL http://www.ncbi.nlm.nih.gov/pubmed/11595186.

Hemminki, A., Peltomäki, P., Mecklin, J. P., Järvinen, H., Salovaara, R., Nyström-Lahti, M., de la Chapelle, A., \& Aaltonen, L. A. (1994). Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. Nature genetics, 8(4), 405-10. URL http://www.ncbi.nlm.nih.gov/pubmed/7894494.

Hewakapuge, S., van Oorschot, R. A. H., Lewandowski, P., \& Baindur-Hudson, S. (2008). Investigation of telomere lengths measurement by quantitative real-time PCR to predict age. Legal medicine (Tokyo, Japan), 10(5), 236-42. URL http: //www.ncbi.nlm.nih.gov/pubmed/18378481.

Hills, M. (2004). The Xp / Yp telomere and its adjacent DNA : structure, sequence organisation and mutation. Ph.D. thesis, Leicester.

Hills, M., Lücke, K., Chavez, E. a., Eaves, C. J., \& Lansdorp, P. M. (2009). Probing the mitotic history and developmental stage of hematopoietic cells using single telomere length analysis (STELA). Blood, 113(23), 5765-75. URL http: //www.ncbi.nlm.nih.gov/pubmed/19359409.

Hiyama, E., Yokoyama, T., Tatsumoto, N., Hiyama, K., Imamura, Y., Murakami, Y., Kodama, T., Piatyszek, M. a., Shay, J. W., \& Matsuura, Y. (1995). Telomerase activity in gastric cancer. Cancer research, 55(15), 3258-62. URL http: //www.ncbi.nlm.nih.gov/pubmed/7614459.

Hjelmborg, J. B., Dalgård, C., Mangino, M., Spector, T. D., Halekoh, U., Möller, S., Kimura, M., Horvath, K., Kark, J. D., Christensen, K., Kyvik, K. O., \& Aviv, A. (2015). Paternal age and telomere length in twins: the germ stem cell selection paradigm. Aging cell, 14(4), 701-3. URL http://www.ncbi.nlm.nih.gov/pubmed/25865872.

Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. Nature, 411(6835), 366-74. URL http://www.ncbi.nlm.nih.gov/pubmed/11357144.

Hohensinner, P. J., Kaun, C., Buchberger, E., Ebenbauer, B., Demyanets, S., Huk, I., Eppel, W., Maurer, G., Huber, K., \& Wojta, J. (2016). Age intrinsic loss of telomere protection via TRF1 reduction in endothelial cells. Biochimica et biophysica acta, 1863(2), 360-7. URL http://www.ncbi.nlm.nih.gov/pubmed/26658719.

Holliday, R. (2007). A mechanism for gene conversion in fungi. Genetical research, 89(5-6), 285-307. URL http://www.ncbi.nlm.nih.gov/pubmed/18976517.

Holohan, B., De Meyer, T., Batten, K., Mangino, M., Hunt, S. C., Bekaert, S., De Buyzere, M. L., Rietzschel, E. R., Spector, T. D., Wright, W. E., \& Shay, J. W. (2015). Decreasing initial telomere length in humans intergenerationally understates age-associated telomere shortening. Aging cell, 14(4), 669-77. URL http: //www.ncbi.nlm.nih.gov/pubmed/25952108.

Honda, S., Hjelmeland, L. M., \& Handa, J. T. (2001). Oxidative stress-induced singlestrand breaks in chromosomal telomeres of human retinal pigment epithelial cells in vitro. Investigative ophthalmology \& visual science, 42(9), 2139-44. URL http: //www.ncbi.nlm.nih.gov/pubmed/11481283.

Houben, J. M. J., Giltay, E. J., Rius-Ottenheim, N., Hageman, G. J., \& Kromhout, D. (2011). Telomere length and mortality in elderly men: the Zutphen Elderly Study. The journals of gerontology. Series A, Biological sciences and medical sciences, 66(1), 38-44. URL http: //www.ncbi.nlm.nih.gov/pubmed/20889650.

Hsieh, P. \& Yamane, K. (2008). DNA mismatch repair: molecular mechanism, cancer, and ageing. Mechanisms of ageing and development, 129(7-8), 391-407. URL http: //www.ncbi.nlm.nih.gov/pubmed/18406444.

Huang, Y., Hidalgo-Bravo, A., Zhang, E., Cotton, V. E., Mendez-Bermudez, A., Wig, G., Medina-Calzada, Z., Neumann, R., Jeffreys, A. J., Winney, B., Wilson, J. F., Clark, D. A., Dyer, M. J., \& Royle, N. J. (2014). Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome. Nucleic acids research, 42(1), 315-27. URL http: //www.ncbi.nlm.nih.gov/pubmed/24057213.

Hunt, S. C., Chen, W., Gardner, J. P., Kimura, M., Srinivasan, S. R., Eckfeldt, J. H., Berenson, G. S., \& Aviv, A. (2008). Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. Aging cell, 7(4), 451-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/18462274.

Hurley, L. H. (2002). DNA and its associated processes as targets for cancer therapy. Nature reviews. Cancer, 2(3), 188-200. URL http: //www.ncbi.nlm.nih.gov/pubmed/11990855.

Hwang, H., Buncher, N., Opresko, P. L., \& Myong, S. (2012). POT1-TPP1 regulates telomeric overhang structural dynamics. Structure (London, England : 1993), 20(11), 1872-80. URL http://www.ncbi.nlm.nih.gov/pubmed/22981946.

Iams, K., Larson, E. D., \& Drummond, J. T. (2002). DNA template requirements for human mismatch repair in vitro. The Journal of biological chemistry, 277(34), 30805-14. URL http://www.ncbi.nlm.nih.gov/pubmed/12077119.

Ip, C. L. C., Loose, M., Tyson, J. R., de Cesare, M., Brown, B. L., Jain, M., Leggett, R. M., Eccles, D. A., Zalunin, V., Urban, J. M., Piazza, P., Bowden, R. J., Paten, B., Mwaigwisya, S., Batty, E. M., Simpson, J. T., Snutch, T. P., Birney, E., Buck, D., Goodwin, S., Jansen, H. J., O'Grady, J., Olsen, H. E., \& MinION Analysis and Reference Consortium (2015). MinION Analysis and Reference Consortium: Phase 1 data release and analysis. F1000Research, 4, 1075. URL http://www.ncbi.nlm.nih.gov/pubmed/26834992.

Iwama, H., Ohyashiki, K., Ohyashiki, J. H., Hayashi, S., Yahata, N., Ando, K., Toyama, K., Hoshika, A., Takasaki, M., Mori, M., \& Shay, J. W. (1998). Telomeric length and telomerase activity vary with age in peripheral blood cells obtained from normal individuals. Human genetics, 102(4), 397-402. URL http://www.ncbi.nlm.nih.gov/pubmed/9600234.

Iyama, T. \& Wilson, D. M. (2013). DNA repair mechanisms in dividing and non-dividing cells. DNA repair, 12(8), 620-36. URL http://www.ncbi.nlm.nih.gov/pubmed/23684800.

Iyer, R. R., Pluciennik, A., Burdett, V., \& Modrich, P. L. (2006). DNA mismatch repair: functions and mechanisms. Chemical reviews, 106(2), 302-23. URL http: //www.ncbi.nlm.nih.gov/pubmed/16464007.

Iyer, R. R., Pluciennik, A., Napierala, M., \& Wells, R. D. (2015). DNA triplet repeat expansion and mismatch repair. Annual review of biochemistry, 84 (December), 199-226. URL http: //www.ncbi.nlm.nih.gov/pubmed/25580529.

Jäger, a. C., Rasmussen, M., Bisgaard, H. C., Singh, K. K., Nielsen, F. C., \& Rasmussen, L. J. (2001). HNPCC mutations in the human DNA mismatch repair gene hMLH1 influence assembly of hMutLalpha and hMLH1-hEXO1 complexes. Oncogene, 20(27), 3590-5. URL http://www.ncbi.nlm.nih.gov/pubmed/11429708.

Jagmohan-Changur, S., Poikonen, T., Vilkki, S., Launonen, V., Wikman, F., Orntoft, T. F., Møller, P., Vasen, H., Tops, C., Kolodner, R. D., Mecklin, J.-p., Järvinen, H., Bevan, S., Houlston, R. S., Aaltonen, L. A., Fodde, R., Wijnen, J., \& Karhu, A. (2003). EXO1 variants occur commonly in normal population: evidence against a role in hereditary nonpolyposis colorectal cancer. Cancer research, 63(1), 154-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/12517792.

Jeffreys, A. J., MacLeod, A., Tamaki, K., Neil, D. L., \& Monckton, D. G. (1991). Minisatellite repeat coding as a digital approach to DNA typing. Nature, 354(6350), 204-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/1961248.

Jeyapalan, J. N., Mendez-Bermudez, A., Zaffaroni, N., Dubrova, Y. E., \& Royle, N. J. (2008). Evidence for alternative lengthening of telomeres in liposarcomas in the absence of ALTassociated PML bodies. International journal of cancer, 122(11), 2414-21. URL http: //www.ncbi.nlm.nih.gov/pubmed/18311780.

Jiang, Y.-A., Luo, H.-S., Zhang, Y.-Y., Fan, L.-F., Jiang, C.-Q., \& Chen, W.-J. (2003). Telomerase activity and cell apoptosis in colon cancer cell by human telomerase reverse transcriptase gene antisense oligodeoxynucleotide. World journal of gastroenterology : WJG, 9(9), 1981-4. URL http://www.ncbi.nlm.nih.gov/pubmed/12970889.

Jiricny, J. (2000). Mismatch repair: the praying hands of fidelity. Current biology : CB, 10(21), R788-90. URL http://www.ncbi.nlm.nih.gov/pubmed/11084351.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nature reviews. Molecular cell biology, 7(5), 335-46. URL http: / /www.ncbi.nlm.nih.gov/pubmed/16612326.

Jones, R. E., Oh, S., Grimstead, J. W., Zimbric, J., Roger, L., Heppel, N. H., Ashelford, K. E., Liddiard, K., Hendrickson, E. A., \& Baird, D. M. (2014). Escape from telomeredriven crisis is DNA ligase III dependent. Cell reports, 8(4), 1063-76. URL http: //www.ncbi.nlm.nih.gov/pubmed/25127141.

Junop, M. S., Obmolova, G., Rausch, K., Hsieh, P., \& Yang, W. (2001). Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. Molecular cell, 7(1), 1-12. URL http: / /www.ncbi.nlm.nih.gov/pubmed/11172706.

Junop, M. S., Yang, W., Funchain, P., Clendenin, W., \& Miller, J. H. (2003). In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. DNA repair, 2(4), 387-405. URL http: //www.ncbi.nlm.nih.gov/pubmed/12606120.

Kadyrov, F. a., Dzantiev, L., Constantin, N., \& Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. Cell, 126(2), 297-308. URL http: //www.ncbi.nlm.nih.gov/pubmed/16873062.

Kalmbach, K., Robinson, L. G., Wang, F., Liu, L., \& Keefe, D. (2014). Telomere length reprogramming in embryos and stem cells. BioMed research international, 2014, 925121. URL http://www.ncbi.nlm.nih.gov/pubmed/24719895.

Kansikas, M., Kasela, M., Kantelinen, J., \& Nyström, M. (2014). Assessing how reduced expression levels of the mismatch repair genes MLH1, MSH2, and MSH6 affect repair efficiency. Human mutation, 35(9), 1123-7. URL http: //www.ncbi.nlm.nih.gov/pubmed/24924810.

Kantelinen, J., Kansikas, M., Candelin, S., Hampel, H., Smith, B., Holm, L., Kariola, R., \& Nyström, M. (2012). Mismatch repair analysis of inherited MSH2 and/or MSH6 variation pairs found in cancer patients. Human mutation, 33(8), 1294-301. URL http: //www.ncbi.nlm.nih.gov/pubmed/22581703.

Kanthan, R., Senger, J.-L., \& Kanthan, S. C. (2012). Molecular events in primary and metastatic colorectal carcinoma: a review. Pathology research international, 2012(since 2006), 597497. URL http://www.ncbi.nlm.nih.gov/pubmed/22997602.

Kariola, R., Raevaara, T. E., Lönnqvist, K. E., \& Nyström-Lahti, M. (2002). Functional analysis of MSH6 mutations linked to kindreds with putative hereditary non-polyposis colorectal cancer syndrome. Human molecular genetics, 11(11), 1303-10. URL http: //www.ncbi.nlm.nih.gov/pubmed/12019211.

Karlseder, J., Smogorzewska, A., \& de Lange, T. (2002). Senescence induced by altered telomere state, not telomere loss. Science (New York, N.Y.), 295(5564), 2446-9. URL http://www.ncbi.nlm.nih.gov/pubmed/11923537.

Karlsson, E., Lärkeryd, A., Sjödin, A., Forsman, M., \& Stenberg, P. (2015). Scaffolding of a bacterial genome using MinION nanopore sequencing. Scientific reports, 5, 11996. URL http://www.ncbi.nlm.nih.gov/pubmed/26149338.

Keefe, D. L. \& Liu, L. (2009). Telomeres and reproductive aging. Reproduction, fertility, and development, 21(1), 10-4. URL http://www.ncbi.nlm.nih.gov/pubmed/19152740.

Kim, J. H. \& Kang, G. H. (2014). Molecular and prognostic heterogeneity of microsatelliteunstable colorectal cancer. World journal of gastroenterology, 20(15), 4230-43. URL http: //www.ncbi.nlm.nih.gov/pubmed/24764661.

Kim, N.-K., Zhang, Q., Zhou, J., Theimer, C. A., Peterson, R. D., \& Feigon, J. (2008). Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. Journal of molecular biology, 384(5), 1249-61. URL http: //www.ncbi.nlm.nih.gov/pubmed/18950640.

Kim, N. W., Piatyszek, M. a., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L., Coviello, G. M., Wright, W. E., Weinrich, S. L., \& Shay, J. W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science (New York, N.Y.), 266(5193), 2011-5. URL http://www.ncbi.nlm.nih.gov/pubmed/7605428.

Kim, S. H., Kaminker, P., \& Campisi, J. (1999). TIN2, a new regulator of telomere length in human cells. Nature genetics, 23(4), 405-12. URL http: //www.ncbi.nlm.nih.gov/pubmed/10581025.

Kim, T.-M. \& Park, P. J. (2014). A Genome-wide View of Microsatellite Instability: Old Stories of Cancer Mutations Revisited with New Sequencing Technologies. Cancer research, 74(22), 6377-6382. URL http://www.ncbi.nlm.nih.gov/pubmed/25371413.

Kimura, M., Cherkas, L. F., Kato, B. S., Demissie, S., Hjelmborg, J. B., Brimacombe, M., Cupples, A., Hunkin, J. L., Gardner, J. P., Lu, X., Cao, X., Sastrasinh, M., Province, M. A., Hunt, S. C., Christensen, K., Levy, D., Spector, T. D., \& Aviv, A. (2008). Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. PLoS genetics, 4(2), e37. URL http://www.ncbi.nlm.nih.gov/pubmed/18282113.

Kimura, M., Stone, R. C., Hunt, S. C., Skurnick, J., Lu, X., Cao, X., Harley, C. B., \& Aviv, A. (2010). Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. Nature protocols, 5(9), 1596-607. URL http: //www.ncbi.nlm.nih.gov/pubmed/21085125.

Kirkpatrick, D. T. (1999). Roles of the DNA mismatch repair and nucleotide excision repair proteins during meiosis. Cellular and molecular life sciences : CMLS, 55(3), 437-49. URL http://www.ncbi.nlm.nih.gov/pubmed/10228557.

Kolas, N. K., Svetlanov, A., Lenzi, M. L., Macaluso, F. P., Lipkin, S. M., Liskay, R. M., Greally, J., Edelmann, W., \& Cohen, P. E. (2005). Localization of MMR proteins on meiotic chromosomes in mice indicates distinct functions during prophase I. The Journal of cell biology, 171(3), 447-58. URL http://www.ncbi.nlm.nih.gov/pubmed/16260499.

Kolodner, R. (1996). Biochemistry and genetics of eukaryotic mismatch repair. Genes \& development, 10(12), 1433-42. URL http://www.ncbi.nlm.nih.gov/pubmed/8666228.

Kolodner, R. D., Tytell, J. D., Schmeits, J. L., Kane, M. F., Gupta, R. D., Weger, J., Wahlberg, S., Fox, E. A., Peel, D., Ziogas, A., Garber, J. E., Syngal, S., Anton-Culver, H., \& Li, F. P. (1999). Germ-line msh6 mutations in colorectal cancer families. Cancer research, 59(20), 5068-74. URL http://www.ncbi.nlm.nih.gov/pubmed/10537275.

Kosar, M., Bartkova, J., Hubackova, S., Hodny, Z., Lukas, J., \& Bartek, J. (2011). Senescenceassociated heterochromatin foci are dispensable for cellular senescence, occur in a cell typeand insult-dependent manner and follow expression of p16(ink4a). Cell cycle (Georgetown, Tex.), 10(3), 457-68. URL http: / /www.ncbi.nlm.nih.gov/pubmed/21248468.

Krokan, H. E., Nilsen, H., Skorpen, F., Otterlei, M., \& Slupphaug, G. (2000). Base excision repair of DNA in mammalian cells. FEBS letters, 476(1-2), 73-7. URL http: //www.ncbi.nlm.nih.gov/pubmed/10878254.

Kunkel, T. a. \& Erie, D. a. (2005). DNA mismatch repair. Annual review of biochemistry, 74, 681-710. URL http://www.ncbi.nlm.nih.gov/pubmed/15952900.

Kuranaga, N., Shinomiya, N., \& Mochizuki, H. (2001). Long-term cultivation of colorectal carcinoma cells with anti-cancer drugs induces drug resistance and telomere elongation: an in vitro study. BMC cancer, 1, 10. URL http: //www.ncbi.nlm.nih.gov/pubmed/11518543.

Lahue, R. S., Au, K. G., \& Modrich, P. (1989). DNA mismatch correction in a defined system. Science (New York, N.Y.), 245(4914), 160-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/2665076.

Lamers, M. H., Perrakis, a., Enzlin, J. H., Winterwerp, H. H., de Wind, N., \& Sixma, T. K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. Nature, 407(6805), 711-7. URL http: //www.ncbi.nlm.nih.gov/pubmed/11048711.

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(5), 685-91. URL http: //www.ncbi.nlm.nih.gov/pubmed/8733138.

Laver, T., Harrison, J., O’Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., \& Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. Biomolecular detection and quantification, 3, 1-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/26753127.

Lever, E. \& Sheer, D. (2010). The role of nuclear organization in cancer. The Journal of pathology, 220(2), 114-25. URL http://www.ncbi.nlm.nih.gov/pubmed/19927301.

Li, B., Oestreich, S., \& de Lange, T. (2000). Identification of human Rap1: implications for telomere evolution. Cell, 101(5), 471-83. URL http: //www.ncbi.nlm.nih.gov/pubmed/10850490.

Li, G.-M. (2008). Mechanisms and functions of DNA mismatch repair. Cell research, 18(1), 85-98. URL http://www.ncbi.nlm.nih.gov/pubmed/18157157.

Li, Y.-C., Korol, A. B., Fahima, T., \& Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. Molecular biology and evolution, 21(6), 991-1007. URL http: //www.ncbi.nlm.nih.gov/pubmed/14963101.

Li, Z. H., Salovaara, R., Aaltonen, L. A., \& Shibata, D. (1996). Telomerase activity is commonly detected in hereditary nonpolyposis colorectal cancers. The American journal of pathology, 148(4), 1075-9. URL http://www.ncbi.nlm.nih.gov/pubmed/8644849.

Liang, F., Han, M., Romanienko, P. J., \& Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America, 95(9), 5172-7. URL http://www.ncbi.nlm.nih.gov/pubmed/9560248.

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(11), 1899-907. URL http://www.ncbi.nlm.nih.gov/pubmed/20538793.

Lin, T. T., Norris, K., Heppel, N. H., Pratt, G., Allan, J. M., Allsup, D. J., Bailey, J., Cawkwell, L., Hills, R., Grimstead, J. W., Jones, R. E., Britt-Compton, B., Fegan, C., Baird, D. M., \& Pepper, C. (2014). Telomere dysfunction accurately predicts clinical outcome in chronic lymphocytic leukaemia, even in patients with early stage disease. British journal of haematology, 167(2), 214-23. URL http: //www.ncbi.nlm.nih.gov/pubmed/24990087.

Lin, Y., Dion, V., \& Wilson, J. H. (2006). Transcription promotes contraction of CAG repeat tracts in human cells. Nature structural \& molecular biology, 13(2), 179-80. URL http: //www.ncbi.nlm.nih.gov/pubmed/16388310.

Lipps, H. J. \& Rhodes, D. (2009). G-quadruplex structures: in vivo evidence and function. Trends in cell biology, 19(8), 414-22. URL http: //www.ncbi.nlm.nih.gov/pubmed/19589679.

Liu, A., Yoshioka, K.-I., Salerno, V., \& Hsieh, P. (2008). The mismatch repair-mediated cell cycle checkpoint response to fluorodeoxyuridine. Journal of cellular biochemistry, 105(1), 245-54. URL http://www.ncbi.nlm.nih.gov/pubmed/18543256.

Liu, A.-Q., Ge, L.-Y., Lu, X.-L., Luo, X.-L., Cai, Y.-L., Ye, X.-Q., \& Geng, F.-F. (2014). Silencing of the hTERT gene by shRNA inhibits colon cancer SW480 cell growth in vitro and in vivo. PloS one, 9(9), e107019. URL http://www.ncbi.nlm.nih.gov/pubmed/25207650.

Liu, B., Nicolaides, N. C., Markowitz, S., Willson, J. K., Parsons, R. E., Jen, J., Papadopolous, N., Peltomäki, P., de la Chapelle, A., \& Hamilton, S. R. (1995). Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. Nature genetics, 9(1), 48-55. URL http://www.ncbi.nlm.nih.gov/pubmed/7704024.

Liu, D., Safari, A., O'Connor, M. S., Chan, D. W., Laegeler, A., Qin, J., \& Songyang, Z. (2004). PTOP interacts with POT1 and regulates its localization to telomeres. Nature cell biology, 6(7), 673-80. URL http://www.ncbi.nlm.nih.gov/pubmed/15181449.

Liu, J.-P. \& Li, H. (2010). Telomerase in the ovary. Reproduction (Cambridge, England), 140(2), 215-22. URL http://www.ncbi.nlm.nih.gov/pubmed/20562297.

Lokanga, R. A., Zhao, X.-N., \& Usdin, K. (2014). The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model. Human mutation, 35(1), 129-36. URL http://www.ncbi.nlm.nih.gov/pubmed/24130133.

Londoño-Vallejo, J. a., DerSarkissian, H., Cazes, L., \& Thomas, G. (2001). Differences in telomere length between homologous chromosomes in humans. Nucleic acids research, 29(15), 3164-71. URL http://www.ncbi.nlm.nih.gov/pubmed/11470873.

Lord, C. J. \& Ashworth, A. (2012). The DNA damage response and cancer therapy. Nature, 481(7381), 287-94. URL http://www.ncbi.nlm.nih.gov/pubmed/22258607.

Lorenz, M., Saretzki, G., Sitte, N., Metzkow, S., \& von Zglinicki, T. (2001). BJ fibroblasts display high antioxidant capacity and slow telomere shortening independent of hTERT transfection. Free radical biology \& medicine, 31(6), 824-31. URL http: //www.ncbi.nlm.nih.gov/pubmed/11557321.

Low, K. C. \& Tergaonkar, V. (2013). Telomerase: central regulator of all of the hallmarks of cancer. Trends in biochemical sciences, 38(9), 426-34. URL http: //www.ncbi.nlm.nih.gov/pubmed/23932019.

Lu, W., Zhang, Y., Liu, D., Songyang, Z., \& Wan, M. (2012). Telomeres-structure, function, and regulation. Experimental cell research, 319(2), 133-141. URL http: //www.ncbi.nlm.nih.gov/pubmed/23006819.

Lujan, S. a., Clark, A. B., \& Kunkel, T. a. (2015). Differences in genome-wide repeat sequence instability conferred by proofreading and mismatch repair defects. Nucleic acids research, 43(8), 4067-74. URL http://www.ncbi.nlm.nih.gov/pubmed/25824945.

Lujan, S. a., Williams, J. S., Pursell, Z. F., Abdulovic-Cui, A. a., Clark, A. B., Nick McElhinny, S. a., \& Kunkel, T. a. (2012). Mismatch repair balances leading and lagging strand DNA replication fidelity. PLoS genetics, 8(10), e1003016. URL http: //www.ncbi.nlm.nih.gov/pubmed/23071460.

Luke-Glaser, S., Poschke, H., \& Luke, B. (2012). Getting in (and out of) the loop: regulating higher order telomere structures. Frontiers in oncology, 2(November), 180. URL http: //www.ncbi.nlm.nih.gov/pubmed/23226680.

Lynch, H. T. \& de la Chapelle, A. (2003). Hereditary colorectal cancer. The New England journal of medicine, 348(10), 919-32. URL http: //www.ncbi.nlm.nih.gov/pubmed/12621137.

Lynch, H. T., Lynch, P. M., Lanspa, S. J., Snyder, C. L., Lynch, J. F., \& Boland, C. R. (2009). Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clinical genetics, 76(1), 1-18. URL http://www.ncbi.nlm.nih.gov/pubmed/22776989.

Ma, H., Zhou, Z., Wei, S., Liu, Z., Pooley, K. a., Dunning, A. M., Svenson, U., Roos, G., Hosgood, H. D., Shen, M., \& Wei, Q. (2011). Shortened telomere length is associated with increased risk of cancer: a meta-analysis. PloS one, 6(6), e20466. URL http://www.ncbi.nlm.nih.gov/pubmed/21695195.

Madoui, M.-A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P., \& Aury, J.-M. (2015). Genome assembly using Nanoporeguided long and error-free DNA reads. BMC genomics, 16, 327. URL http: //www.ncbi.nlm.nih.gov/pubmed/25927464.

Manley, K., Shirley, T. L., Flaherty, L., \& Messer, A. (1999). Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. Nature genetics, 23(4), 471-3. URL http://www.ncbi.nlm.nih.gov/pubmed/10581038.

Marteijn, J. A., Lans, H., Vermeulen, W., \& Hoeijmakers, J. H. J. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. Nature reviews. Molecular cell biology, 15(7), 465-81. URL http://www.ncbi.nlm.nih.gov/pubmed/24954209.

Martens, U. M., Brass, V., Sedlacek, L., Pantic, M., Exner, C., Guo, Y., Engelhardt, M., Lansdorp, P. M., Waller, C. F., \& Lange, W. (2002). Telomere maintenance in human B lymphocytes. British journal of haematology, 119(3), 810-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/12437664.

Marthandan, S., Priebe, S., Hemmerich, P., Klement, K., \& Diekmann, S. (2014). Longterm quiescent fibroblast cells transit into senescence. PloS one, 9(12), e115597. URL http://www.ncbi.nlm.nih.gov/pubmed/25531649.

Martin, S. a., McCabe, N., Mullarkey, M., Cummins, R., Burgess, D. J., Nakabeppu, Y., Oka, S., Kay, E., Lord, C. J., \& Ashworth, A. (2010). DNA polymerases as potential therapeutic targets for cancers deficient in the DNA mismatch repair proteins MSH2 or MLH1. Cancer cell, 17(3), 235-48. URL http: / /www.ncbi.nlm.nih.gov/pubmed/20227038.

Martin-Ruiz, C. M., Baird, D., Roger, L., Boukamp, P., Krunic, D., Cawthon, R., Dokter, M. M., van der Harst, P., Bekaert, S., de Meyer, T., Roos, G., Svenson, U., Codd, V., Samani, N. J., McGlynn, L., Shiels, P. G., Pooley, K. A., Dunning, A. M., Cooper, R., Wong, A., Kingston, A., \& von Zglinicki, T. (2015). Reproducibility of Telomere Length AssessmentAn International Collaborative Study. International journal of epidemiology, 44(5), 174954. URL http://www.ncbi.nlm.nih.gov/pubmed/26403809.

Martin-Ruiz, C. M., Gussekloo, J., van Heemst, D., von Zglinicki, T., \& Westendorp, R. G. J. (2005). Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. Aging cell, 4(6), 287-90. URL http://www.ncbi.nlm.nih.gov/pubmed/16300480.

Martínez, P. \& Blasco, M. a. (2011). Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. Nature reviews. Cancer, 11(3), 161-76. URL http: //www.ncbi.nlm.nih.gov/pubmed/21346783.

Martinez, P., Siegl-Cachedenier, I., Flores, J. M., \& Blasco, M. a. (2009). MSH2 deficiency abolishes the anticancer and pro-aging activity of short telomeres. Aging cell, 8(1), 2-17. URL http://www.ncbi.nlm.nih.gov/pubmed/18986375.

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(7), e1002182. URL http: //www.ncbi.nlm.nih.gov/pubmed/21829373.

Mastrocola, A. S. \& Heinen, C. D. (2010). Lynch syndrome-associated mutations in MSH2 alter DNA repair and checkpoint response functions in vivo. Human mutation, 31(10), E1699-708. URL http://www.ncbi.nlm.nih.gov/pubmed/20672385.

McCulloch, S. D. \& Kunkel, T. a. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell research, 18(1), 148-61. URL http://www.ncbi.nlm.nih.gov/pubmed/18166979.

Mendez-Bermudez, A., Hidalgo-Bravo, A., Cotton, V. E., Gravani, A., Jeyapalan, J. N., \& Royle, N. J. (2012). The roles of WRN and BLM RecQ helicases in the Alternative Lengthening of Telomeres. Nucleic acids research, 40(21), 10809-20. URL http://www.ncbi.nlm.nih.gov/pubmed/22989712http: //www. pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3510502.

Mendez-Bermudez, A., Hills, M., Pickett, H. a., Phan, A. T., Mergny, J.-L., Riou, J.-F., \& Royle, N. J. (2009). Human telomeres that contain (CTAGGG)n repeats show replication dependent instability in somatic cells and the male germline. Nucleic acids research, 37(18), 6225-38. URL http://www.ncbi.nlm.nih.gov/pubmed/19656953.

Mendez-Bermudez, A. \& Royle, N. J. (2011). Deficiency in DNA mismatch repair increases the rate of telomere shortening in normal human cells. Human mutation, 32(8), 939-46. URL http://www.ncbi.nlm.nih.gov/pubmed/21538690.

Mikheyev, A. S. \& Tin, M. M. Y. (2014). A first look at the Oxford Nanopore MinION sequencer. Molecular ecology resources, 14(6), 1097-102. URL http: //www.ncbi.nlm.nih.gov/pubmed/25187008.

Mitchell, C., Hobcraft, J., McLanahan, S. S., Siegel, S. R., Berg, A., BrooksGunn, J., Garfinkel, I., \& Notterman, D. (2014). Social disadvantage, genetic sensitivity, and children's telomere length. Proceedings of the National Academy of Sciences of the United States of America, 111(16), 5944-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/24711381.

Modrich, P. (1987). DNA mismatch correction. Annual review of biochemistry, 56(12), 435-66. URL http://www.ncbi.nlm.nih.gov/pubmed/2144613.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. Annual review of genetics, 25, 229-53. URL http://www.ncbi.nlm.nih.gov/pubmed/1812808.

Modrich, P. \& Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annual review of biochemistry, 65, 101-33. URL http: //www.ncbi.nlm.nih.gov/pubmed/8811176.

Mondello, C., Petropoulou, C., Monti, D., Gonos, E. S., Franceschi, C., \& Nuzzo, F. (1999). Telomere length in fibroblasts and blood cells from healthy centenarians. Experimental cell research, 248(1), 234-42. URL http://www.ncbi.nlm.nih.gov/pubmed/10094830.

Montpetit, A. J., Alhareeri, A. A., Montpetit, M., Starkweather, A. R., Elmore, L. W., Filler, K., Mohanraj, L., Burton, C. W., Menzies, V. S., Lyon, D. E., \& Jackson-Cook, C. K. (2015). Telomere length: a review of methods for measurement. Nursing research, 63(4), 289-99. URL http://www.ncbi.nlm.nih.gov/pubmed/24977726.

Morales, F., Vásquez, M., Santamaría, C., Cuenca, P., Corrales, E., \& Monckton, D. G. (2016). A polymorphism in the MSH3 mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients. DNA repair, 40, 57-66. URL http: //www.ncbi.nlm.nih.gov/pubmed/26994442.

Muller, H. J. \& Altenburg, E. (1930). The Frequency of Translocations Produced by X-Rays in Drosophila. Genetics, 15(4), 283-311. URL http: //www.ncbi.nlm.nih.gov/pubmed/17246601.

Muñoz-Espín, D. \& Serrano, M. (2014). Cellular senescence: from physiology to pathology. Nature reviews. Molecular cell biology, 15(7), 482-96. URL http: //www.ncbi.nlm.nih.gov/pubmed/24954210.

Munro, J., Steeghs, K., Morrison, V., Ireland, H., \& Parkinson, E. K. (2001). Human fibroblast replicative senescence can occur in the absence of extensive cell division and short telomeres. Oncogene, 20(27), 3541-52. URL http: //www.ncbi.nlm.nih.gov/pubmed/11429701.

Muraki, K., Nyhan, K., Han, L., \& Murnane, J. P. (2012). Mechanisms of telomere loss and their consequences for chromosome instability. Frontiers in oncology, 2(October), 135. URL http://www.ncbi.nlm.nih.gov/pubmed/23061048.

Nakamura, K., Furugori, E., Esaki, Y., Arai, T., Sawabe, M., Okayasu, I., Fujiwara, M., Kammori, M., Mafune, K., Kato, M., Oshimura, M., Sasajima, K., \& Takubo, K. (2000). Correlation of telomere lengths in normal and cancers tissue in the large bowel. Cancer letters, 158(2), 179-84. URL http://www.ncbi.nlm.nih.gov/pubmed/10960768.

Nakamura, K.-I., Izumiyama-Shimomura, N., Sawabe, M., Arai, T., Aoyagi, Y., Fujiwara, M., Tsuchiya, E., Kobayashi, Y., Kato, M., Oshimura, M., Sasajima, K., Nakachi, K., \& Takubo, K. (2002). Comparative analysis of telomere lengths and erosion with age in human epidermis and lingual epithelium. The Journal of investigative dermatology, 119(5), 1014-9. URL http://www.ncbi.nlm.nih.gov/pulomed/12445186.

Nakashima, M., Nandakumar, J., Sullivan, K. D., Espinosa, J. M., \& Cech, T. R. (2013). Inhibition of telomerase recruitment and cancer cell death. The Journal of biological chemistry, 288(46), 33171-80. URL http://www.ncbi.nlm.nih.gov/pubmed/24097987.

Nakayama, A., Sato, M., Shinohara, M., Matsubara, S., Yokomine, T., Akasaka, E., Yoshida, M., \& Takao, S. (2007). Efficient transfection of primarily cultured porcine embryonic fibroblasts using the Amaxa Nucleofection system. Cloning and stem cells, 9(4), 523-34. URL http://www.ncbi.nlm.nih.gov/pubmed/18154513.

Nandakumar, J. \& Cech, T. R. (2013). Finding the end: recruitment of telomerase to telomeres. Nature reviews. Molecular cell biology, 14(2), 69-82. URL http: //www.ncbi.nlm.nih.gov/pubmed/23299958.

Nawrot, T. S., Staessen, J. a., Gardner, J. P., \& Aviv, A. (2004). Telomere length and possible link to X chromosome. Lancet (London, England), 363(9408), 507-10. URL http://www.ncbi.nlm.nih.gov/pubmed/14975611.

Nelson, D. M., McBryan, T., Jeyapalan, J. C., Sedivy, J. M., \& Adams, P. D. (2014). A comparison of oncogene-induced senescence and replicative senescence: implications for tumor suppression and aging. Age (Dordrecht, Netherlands), 36(3), 9637. URL http: //www.ncbi.nlm.nih.gov/pubmed/24647599.

Nera, B., Huang, H.-S., Lai, T., \& Xu, L. (2015). Elevated levels of TRF2 induce telomeric ultrafine anaphase bridges and rapid telomere deletions. Nature communications, $6,10132$. URL http://www.ncbi.nlm.nih.gov/pubmed/26640040.

New, L., Liu, K., \& Crouse, G. F. (1993). The yeast gene MSH3 defines a new class of eukaryotic MutS homologues. Molecular \& general genetics : MGG, 239(1-2), 97-108. URL http://www.ncbi.nlm.nih.gov/pubmed/8510668.

Neyton, S., Lespinasse, F., Moens, P. B., Paul, R., Gaudray, P., Paquis-Flucklinger, V., \& Santucci-Darmanin, S. (2004). Association between MSH4 (MutS homologue 4) and the DNA strand-exchange RAD51 and DMC1 proteins during mammalian meiosis. Molecular human reproduction, 10(12), 917-24. URL http: //www.ncbi.nlm.nih.gov/pubmed/15489243.

Nguyen, B. N., Elmore, L. W., \& Holt, S. E. (2009). Mechanism of dominant-negative telomerase function. Cell cycle (Georgetown, Tex.), 8(19), 3227-33. URL http: //www.ncbi.nlm.nih.gov/pubmed/19738429.

Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., \& Fraser, C. M. (1994). Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature, 371(6492), 75-80. URL http://www.ncbi.nlm.nih.gov/pubmed/8072530.

Nilbert, M., Timshel, S., Bernstein, I., \& Larsen, K. (2009). Role for genetic anticipation in Lynch syndrome. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 27(3), 360-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/19075283.

Nishant, K. T., Singh, N. D., \& Alani, E. (2009). Genomic mutation rates: what high-throughput methods can tell us. BioEssays: news and reviews in molecular, cellular and developmental biology, 31(9), 912-20. URL http: //www.ncbi.nlm.nih.gov/pubmed/19644920.

Njajou, O. T., Cawthon, R. M., Damcott, C. M., Wu, S.-H., Ott, S., Garant, M. J., Blackburn, E. H., Mitchell, B. D., Shuldiner, A. R., \& Hsueh, W.-C. (2007). Telomere length is paternally inherited and is associated with parental lifespan. Proceedings of the National Academy of Sciences of the United States of America, 104(29), 12135-9. URL http://www.ncbi.nlm.nih.gov/pubmed/17623782.

Nordfjäll, K., Larefalk, A., Lindgren, P., Holmberg, D., \& Roos, G. (2005). Telomere length and heredity: Indications of paternal inheritance. Proceedings of the National Academy of Sciences of the United States of America, 102(45), 16374-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/16258070.

Nordfjäll, K., Svenson, U., Norrback, K.-F., Adolfsson, R., \& Roos, G. (2010). Largescale parent-child comparison confirms a strong paternal influence on telomere length. European journal of human genetics : EJHG, 18(3), 385-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/19826452.

Obmolova, G., Ban, C., Hsieh, P., \& Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature, 407(6805), 703-10. URL http://www.ncbi.nlm.nih.gov/pubmed/11048710.

O’Donnell, M., Kuriyan, J., Kong, X. P., Stukenberg, P. T., \& Onrust, R. (1992). The sliding clamp of dna polymerase iii holoenzyme encircles dna. Mol Biol Cell, 3(9), 953-7.

Olivero, O. A., Anderson, L. M., Diwan, B. A., Haines, D. C., Harbaugh, S. W., Moskal, T. J., Jones, A. B., Rice, J. M., Riggs, C. W., Logsdon, D., Yuspa, S. H., \& Poirier, M. C. (1997). Transplacental effects of 3'-azido-2',3'-dideoxythymidine (AZT): tumorigenicity in mice and genotoxicity in mice and monkeys. Journal of the National Cancer Institute, 89(21), 16028. URL http://www.ncbi.nlm.nih.gov/pubmed/9362158.

Ollila, S., Dermadi Bebek, D., Jiricny, J., \& Nyström, M. (2008). Mechanisms of pathogenicity in human MSH2 missense mutants. Human mutation, 29(11), 1355-63. URL http: //www.ncbi.nlm.nih.gov/pubmed/18951462.

Opresko, P. L., von Kobbe, C., Laine, J.-P., Harrigan, J., Hickson, I. D., \& Bohr, V. a. (2002). Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. The Journal of biological chemistry, 277(43), 41110-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/12181313.

O'Sullivan, R. J. \& Karlseder, J. (2010). Telomeres: protecting chromosomes against genome instability. Nat Rev Mol Cell Biol, 11(3), 171-81.

Ouellette, M. M., Liao, M., Herbert, B. S., Johnson, M., Holt, S. E., Liss, H. S., Shay, J. W., \& Wright, W. E. (2000). Subsenescent telomere lengths in fibroblasts immortalized by limiting amounts of telomerase. The Journal of biological chemistry, 275(14), 10072-6. URL http://www.ncbi.nlm.nih.gov/pubmed/10744686.

Ozaslan, M. \& Aytekin, T. (2009). Loss of heterozygosity in colorectal cancer. African Journal of Bitechnology, 8(25), 7308-7312. URL http: //www.academicjournals.org/journal/AJB/article-abstract/127B6D032438.

Ozturk, S. (2015). Telomerase activity and telomere length in male germ cells. Biology of reproduction, 92(2), 53. URL http://www.ncbi.nlm.nih.gov/pubmed/25568305.

Ozturk, S., Sozen, B., \& Demir, N. (2014). Telomere length and telomerase activity during oocyte maturation and early embryo development in mammalian species. Molecular human reproduction, 20(1), 15-30. URL http: //www.ncbi.nlm.nih.gov/pubmed/23928157.

Pabla, N., Ma, Z., McIlhatton, M. a., Fishel, R., \& Dong, Z. (2011). hMSH2 recruits ATR to DNA damage sites for activation during DNA damage-induced apoptosis. The Journal of biological chemistry, 286(12), 10411-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/21285353.

Palm, W. \& de Lange, T. (2008). How shelterin protects mammalian telomeres. Annual review of genetics, 42, 301-34. URL http: //www.ncbi.nlm.nih.gov/pubmed/18680434.

Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D’Arrigo, A., Truong, O., Hsuan, J. J., \& Jiricny, J. (1995). GTBP, a 160-kilodalton protein essential for mismatchbinding activity in human cells. Science (New York, N.Y.), 268(5219), 1912-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/7604265.

Pang, P. P., Lundberg, a. S., \& Walker, G. C. (1985). Identification and characterization of the mutL and mutS gene products of Salmonella typhimurium LT2. Journal of bacteriology, 163(3), 1007-15. URL http://www.ncbi.nlm.nih.gov/pubmed/2993227.

Panigrahi, G. B., Slean, M. M., Simard, J. P., \& Pearson, C. E. (2012). Human mismatch repair protein $\mathrm{hMutL} \alpha$ is required to repair short slipped-DNAs of trinucleotide repeats. The Journal of biological chemistry, 287(50), 41844-50. URL http: //www.ncbi.nlm.nih.gov/pubmed/23086927.

Papadopoulos, N. \& Lindblom, A. (1997). Molecular basis of HNPCC: mutations of MMR genes. Human mutation, 10(2), 89-99. URL http: //www.ncbi.nlm.nih.gov/pubmed/9259192.

Papadopoulos, N., Nicolaides, N. C., Liu, B., Parsons, R., Lengauer, C., Palombo, F., D'Arrigo, A., Markowitz, S., Willson, J. K., \& Kinzler, K. W. (1995). Mutations of GTBP in genetically unstable cells. Science (New York, N.Y.), 268(5219), 1915-7. URL http: //www.ncbi.nlm.nih.gov/pubmed/7604266.

Papadopoulos, N., Nicolaides, N. C., Wei, Y. F., Ruben, S. M., Carter, K. C., Rosen, C. a., Haseltine, W. a., Fleischmann, R. D., Fraser, C. M., \& Adams, M. D. (1994). Mutation of a mutL homolog in hereditary colon cancer. Science (New York, N.Y.), 263(5153), 1625-9. URL http://www.ncbi.nlm.nih.gov/pubmed/8128251.

Parsch, D., Brassat, U., Brümmendorf, T. H., \& Fellenberg, J. (2008). Consequences of telomerase inhibition by BIBR1532 on proliferation and chemosensitivity of chondrosarcoma cell lines. Cancer investigation, 26(6), 590-6. URL http: //www.ncbi.nlm.nih.gov/pubmed/18584350.

Parsons, R., Li, G. M., Longley, M., Modrich, P., Liu, B., Berk, T., Hamilton, S. R., Kinzler, K. W., \& Vogelstein, B. (1995). Mismatch repair deficiency in phenotypically normal human cells. Science (New York, N.Y.), 268(5211), 738-40. URL http: //www.ncbi.nlm.nih.gov/pubmed/7632227.

Pascolo, E., Wenz, C., Lingner, J., Hauel, N., Priepke, H., Kauffmann, I., Garin-Chesa, P., Rettig, W. J., Damm, K., \& Schnapp, A. (2002). Mechanism of human telomerase inhibition by BIBR1532, a synthetic, non-nucleosidic drug candidate. The Journal of biological chemistry, 277(18), 15566-72. URL http://www.ncbi.nlm.nih.gov/pubmed/11854300.

Peltomäki, P. (2016). Update on Lynch syndrome genomics. Familial cancer. URL http: //www.ncbi.nlm.nih.gov/pubmed/26873718.

Peltomäki, P., Aaltonen, L. a., Sistonen, P., Pylkkänen, L., Mecklin, J. P., Järvinen, H., Green, J. S., Jass, J. R., Weber, J. L., \& Leach, F. S. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. Science (New York, N.Y.), 260(5109), 810-2. URL http://www.ncbi.nlm.nih.gov/pubmed/8484120.

Peltomäki, P. \& Vasen, H. F. (1997). Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. Gastroenterology, 113(4), 1146-58. URL http://www.ncbi.nlm.nih.gov/pubmed/9322509.

Petersen, S., Saretzki, G., \& von Zglinicki, T. (1998). Preferential accumulation of singlestranded regions in telomeres of human fibroblasts. Experimental cell research, 239(1), 152-60. URL http://www.ncbi.nlm.nih.gov/pubmed/9511733.

Phatak, P. \& Burger, a. M. (2007). Telomerase and its potential for therapeutic intervention. British journal of pharmacology, 152(7), 1003-11. URL http: //www.ncbi.nlm.nih.gov/pubmed/17603541.

Piatyszek, M. a., Kim, N. W., Weinrich, S. L., Hiyama, K., Hiyama, E., Wright, W. E., \& Shay, J. W. (1995). Detection of telomerase activity in human cells and tumors by a telomeric repeat amplification protocol (TRAP). Methods in Cell Science, 17(1), 1-15. URL http: //link.springer.com/10.1007/BF00981880.

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(19), 3434-43. URL http: //www.ncbi.nlm.nih.gov/pubmed/15048084.

Pickett, H. a., Henson, J. D., Au, A. Y. M., Neumann, A. a., \& Reddel, R. R. (2011). Normal mammalian cells negatively regulate telomere length by telomere trimming. Human molecular genetics, 20(23), 4684-92. URL http: //www.ncbi.nlm.nih.gov/pubmed/21903669.

Pillon, M. C., Miller, J. H., \& Guarné, A. (2011). The endonuclease domain of MutL interacts with the $\beta$ sliding clamp. DNA repair, 10(1), 87-93. URL http: //www.ncbi.nlm.nih.gov/pubmed/21050827.

Pino, M. S. \& Chung, D. C. (2010). The chromosomal instability pathway in colon cancer. Gastroenterology, 138(6), 2059-72. URL http: //www.ncbi.nlm.nih.gov/pubmed/20420946.

Plumb, J. a., Bilsland, a., Kakani, R., Zhao, J., Glasspool, R. M., Knox, R. J., Evans, T. R., \& Keith, W. N. (2001). Telomerase-specific suicide gene therapy vectors expressing bacterial nitroreductase sensitize human cancer cells to the pro-drug CB1954. Oncogene, 20(53), 7797-803. URL http://www.ncbi.nlm.nih.gov/pubmed/12545158.

Podlevsky, J. D. \& Chen, J. J.-L. (2012). It all comes together at the ends: telomerase structure, function, and biogenesis. Mutation research, 730(1-2), 3-11. URL http: //www.ncbi.nlm.nih.gov/pubmed/22093366.

Pooley, K. a., McGuffog, L., Barrowdale, D., Frost, D., Ellis, S. D., Fineberg, E., Platte, R., Izatt, L., Adlard, J., Bardwell, J., Brewer, C., Cole, T., Cook, J., Davidson, R., Donaldson, A., Dorkins, H., Douglas, F., Eason, J., Houghton, C., Kennedy, M. J., McCann, E., Miedzybrodzka, Z., Murray, A., Porteous, M. E., Rogers, M. T., Side, L. E., Tischkowitz, M., Walker, L., Hodgeson, S., Eccles, D. M., Morrison, P. J., Evans, D. G. R., Eeles, R., Antoniou, A. C., Easton, D. F., \& Dunning, A. M. (2014). Lymphocyte telomere length is longer in BRCA1 and BRCA2 mutation carriers but does not affect subsequent cancer risk. Cancer epidemiology, biomarkers \& prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. URL http://www.ncbi.nlm.nih.gov/pubmed/24642354.

Pope, C. F., O'Sullivan, D. M., McHugh, T. D., \& Gillespie, S. H. (2008). A practical guide to measuring mutation rates in antibiotic resistance.

Prats, H., Martin, B., \& Claverys, J. P. (1985). The hexB mismatch repair gene of Streptococcus pneumoniae: characterisation, cloning and identification of the product. Molecular \& general genetics : MGG, 200(3), 482-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/2995767.

Prolla, T. A., Baker, S. M., Harris, A. C., Tsao, J. L., Yao, X., Bronner, C. E., Zheng, B., Gordon, M., Reneker, J., Arnheim, N., Shibata, D., Bradley, A., \& Liskay, R. M. (1998). Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. Nature genetics, 18(3), 276-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/9500552.

Prolla, T. A., Christie, D. M., \& Liskay, R. M. (1994). Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene. Molecular and cellular biology, 14(1), 407-15. URL http: //www.ncbi.nlm.nih.gov/pubmed/8264608.

Pumpernik, D., Oblak, B., \& Borstnik, B. (2008). Replication slippage versus point mutation rates in short tandem repeats of the human genome. Molecular genetics and genomics : MGG, 279(1), 53-61. URL http: / /www.ncbi.nlm.nih.gov/pu.bmed/17926066.

Qin, Q., Sun, J., Yin, J., Liu, L., Chen, J., Zhang, Y., Li, T., Shi, Y., Wei, S., \& Nie, S. (2014). Telomere length in peripheral blood leukocytes is associated with risk of colorectal cancer in Chinese population. PloS one, 9(2), e88135. URL http: //www.ncbi.nlm.nih.gov/pubmed/24498432.

Quick, J., Quinlan, A. R., \& Loman, N. J. (2015). Erratum: A reference bacterial genome dataset generated on the MinION(TM) portable single-molecule nanopore sequencer. GigaScience, 4(1), 6. URL http://www.ncbi.nlm.nih.gov/pubmed/25695306.

Rampazzo, E., Bertorelle, R., Serra, L., Terrin, L., Candiotto, C., Pucciarelli, S., Del Bianco, P., Nitti, D., \& De Rossi, A. (2010). Relationship between telomere shortening, genetic instability, and site of tumour origin in colorectal cancers. British journal of cancer, 102(8), 1300-5. URL http://www.ncbi.nlm.nih.gov/pubmed/20386541.

Rankin, A. M., Faller, D. V., \& Spanjaard, R. a. (2008). Telomerase inhibitors and 'T-oligo' as cancer therapeutics: contrasting molecular mechanisms of cytotoxicity. Anti-cancer drugs, 19(4), 329-38. URL http://www.ncbi.nlm.nih.gov/pubmed/18454043.

Reenan, R. A. \& Kolodner, R. D. (1992). Isolation and characterization of two Saccharomyces cerevisiae genes encoding homologs of the bacterial HexA and MutS mismatch repair proteins. Genetics, 132(4), 963-73. URL http: //www.ncbi.nlm.nih.gov/pubmed/1459447.

Reig-Viader, R., Vila-Cejudo, M., Vitelli, V., Buscà, R., Sabaté, M., Giulotto, E., Garcia Caldés, M., \& Ruiz-Herrera, A. (2014). Telomeric Repeat-Containing RNA (TERRA) and Telomerase Are Components of Telomeres During Mammalian Gametogenesis. Biology of reproduction. URL http://www.ncbi.nlm.nih.gov/pubmed/24719256.

Richards, J. B., Valdes, A. M., Gardner, J. P., Kato, B. S., Siva, A., Kimura, M., Lu, X., Brown, M. J., Aviv, A., \& Spector, T. D. (2008). Homocysteine levels and leukocyte telomere length. Atherosclerosis, 200(2), 271-7. URL http: //www.ncbi.nlm.nih.gov/pubmed/18280483.

Richter, T., Saretzki, G., Nelson, G., Melcher, M., Olijslagers, S., \& von Zglinicki, T. (2007). TRF2 overexpression diminishes repair of telomeric single-strand breaks and accelerates telomere shortening in human fibroblasts. Mechanisms of ageing and development, 128(4), 340-5. URL http://www.ncbi.nlm.nih.gov/pubmed/17395247.

Risinger, J. I., Umar, A., Boyd, J., Berchuck, A., Kunkel, T. A., \& Barrett, J. C. (1996). Mutation of MSH3 in endometrial cancer and evidence for its functional role in heteroduplex repair. Nature genetics, 14(1), 102-5. URL http: //www.ncbi.nlm.nih.gov/pubmed/8782829.

Rizki, a. \& Lundblad, V. (2001). Defects in mismatch repair promote telomerase-independent proliferation. Nature, 411(6838), 713-6. URL http: //www.ncbi.nlm.nih.gov/pubmed/11395777.

Rouda, S. \& Skordalakes, E. (2007). Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. Structure (London, England : 1993), 15(11), 1403-12. URL http://www.ncbi.nlm.nih.gov/pubmed/17997966.

Rufer, N., Brümmendorf, T. H., Kolvraa, S., Bischoff, C., Christensen, K., Wadsworth, L., Schulzer, M., \& Lansdorp, P. M. (1999). Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. The Journal of experimental medicine, 190(2), 157-67. URL http://www.ncbi.nlm.nih.gov/pubmed/10432279.

Rufer, N., Dragowska, W., Thornbury, G., Roosnek, E., \& Lansdorp, P. M. (1998). Telomere length dynamics in human lymphocyte subpopulations measured by flow cytometry. Nature biotechnology, 16(8), 743-7. URL http: //www.ncbi.nlm.nih.gov/pubmed/9702772.

Rüschoff, J., Wallinger, S., Dietmaier, W., Bocker, T., Brockhoff, G., Hofstädter, F., \& Fishel, R. (1998). Aspirin suppresses the mutator phenotype associated with hereditary nonpolyposis colorectal cancer by genetic selection. Proceedings of the National Academy of Sciences of the United States of America, 95(19), 11301-6. URL http: //www.ncbi.nlm.nih.gov/pubmed/9736731.

Rust, M. J., Bates, M., \& Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nature methods, 3(10), 793-5. URL http: //www.ncbi.nlm.nih.gov/pubmed/16896339.

Rustgi, A. K. (2007). The genetics of hereditary colon cancer. Genes \& development, 21(20), 2525-38. URL http://www.ncbi.nlm.nih.gov/pubmed/17938238.

Samy, M., Gattolliat, C.-H., Pendino, F., Hillion, J., Nguyen, E., Bombard, S., DoucRasy, S., Bénard, J., \& Ségal-Bendirdjian, E. (2012). Loss of the malignant phenotype of human neuroblastoma cells by a catalytically inactive dominant-negative hTERT mutant. Molecular cancer therapeutics, 11(11), 2384-93. URL http: //www.ncbi.nlm.nih.gov/pubmed/22933702.

Sanger, F., Nicklen, S., \& Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America, 74(12), 5463-7. URL http://www.ncbi.nlm.nih.gov/pubmed/271968.

Savouret, C., Garcia-Cordier, C., Megret, J., te Riele, H., Junien, C., \& Gourdon, G. (2004). MSH2-dependent germinal CTG repeat expansions are produced continuously in spermatogonia from DM1 transgenic mice. Mol Cell Biol, 24(2), 629-637. URL http: //www.ncbi.nlm.nih.gov/pubmed/14701736.

Schaetzlein, S., Lucas-Hahn, A., Lemme, E., Kues, W. a., Dorsch, M., Manns, M. P., Niemann, H., \& Rudolph, K. L. (2004). Telomere length is reset during early mammalian embryogenesis. Proceedings of the National Academy of Sciences of the United States of America, 101(21), 8034-8. URL http://www.ncbi.nlm.nih.gov/pubmed/15148368.

Schmidt, M. H. M. \& Pearson, C. E. (2016). Disease-associated repeat instability and mismatch repair. DNA repair, 38, 117-26. URL http: //www.ncbi.nlm.nih.gov/pubmed/26774442.

Schulz, L. \& Tyler, J. (2005). Heterochromatin focuses on senescence. Molecular cell, 17(2), 168-70. URL http://www.ncbi.nlm.nih.gov/pubmed/15664186.

Seguí, N., Guinó, E., Pineda, M., Navarro, M., Bellido, F., Lázaro, C., Blanco, I., Moreno, V., Capellá, G., \& Valle, L. (2014). Longer telomeres are associated with cancer risk in MMR-proficient hereditary non-polyposis colorectal cancer. PloS one, 9(2), e86063. URL http://www.ncbi.nlm.nih.gov/pubmed/24498269.

Seguí, N., Pineda, M., Guinó, E., Borràs, E., Navarro, M., Bellido, F., Moreno, V., Lázaro, C., Blanco, I., Capellá, G., \& Valle, L. (2013). Telomere length and genetic anticipation in Lynch syndrome. PloS one, 8(4), e61286. URL http: //www.ncbi.nlm.nih.gov/pubmed/23637804.

Serra, V. \& von Zglinicki, T. (2002). Human fibroblasts in vitro senesce with a donor-specific telomere length. FEBS letters, 516(1-3), 71-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/17604920.

Serra, V., von Zglinicki, T., Lorenz, M., \& Saretzki, G. (2003). Extracellular superoxide dismutase is a major antioxidant in human fibroblasts and slows telomere shortening. The Journal of biological chemistry, 278(9), 6824-30. URL http: //www.ncbi.nlm.nih.gov/pu.bmed/12475988.

Serrano, M., Lage, P., Belga, S., Filipe, B., Francisco, I., Rodrigues, P., Fonseca, R., Chaves, P., Claro, I., Albuquerque, C., \& Pereira, A. D. (2012). Bethesda criteria for microsatellite instability testing: impact on the detection of new cases of Lynch syndrome. Familial cancer. URL http://www.ncbi.nlm.nih.gov/pubmed/22776989.

Shah, S. N. \& Eckert, K. A. (2009). Human postmeiotic segregation 2 exhibits biased repair at tetranucleotide microsatellite sequences. Cancer research, 69(3), 1143-9. URL http: //www.ncbi.nlm.nih.gov/pubmed/19155293.

Shah, S. N., Hile, S. E., \& Eckert, K. a. (2010). Defective mismatch repair, microsatellite mutation bias, and variability in clinical cancer phenotypes. Cancer research, 70(2), 4315. URL http://www.ncbi.nlm.nih.gov/pubmed/20068152.

Shay, J. W. \& Bacchetti, S. (1997). A survey of telomerase activity in human cancer. European journal of cancer (Oxford, England : 1990), 33(5), 787-91. URL http: //www.ncbi.nlm.nih.gov/pubmed/9282118.

Shibata, D., Peinado, M. a., Ionov, Y., Malkhosyan, S., \& Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. Nature genetics, 6(3), 273-81. URL http: //www.ncbi.nlm.nih.gov/pubmed/8012390.

Sipos, F., Valcz, G., \& Molnár, B. (2012). Physiological and pathological role of local and immigrating colonic stem cells. World journal of gastroenterology, 18(4), 295-301. URL http://www.ncbi.nlm.nih.gov/pubmed/22294835.

Sitte, N., Saretzki, G., \& von Zglinicki, T. (1998). Accelerated telomere shortening in fibroblasts after extended periods of confluency. Free radical biology \& medicine, 24(6), 885-93. URL http://www.ncbi.nlm.nih.gov/pubmed/9607597.

Slagboom, P. E., Droog, S., \& Boomsma, D. I. (1994). Genetic determination of telomere size in humans: a twin study of three age groups. American journal of human genetics, 55(5), 876-82. URL http://www.ncbi.nlm.nih.gov/pubmed/7977349.

Smith, M. J., Urquhart, J. E., Harkness, E. F., Miles, E. K., Bowers, N. L., Byers, H. J., Bulman, M., Gokhale, C., Wallace, A. J., Newman, W. G., \& Evans, D. G. (2016). The Contribution of Whole Gene Deletions and Large Rearrangements to the Mutation Spectrum in Inherited Tumor Predisposing Syndromes. Human mutation, 37(3), 250-6. URL http://www.ncbi.nlm.nih.gov/pubmed/26615784.

Smith, S., Giriat, I., Schmitt, A., \& de Lange, T. (1998). Tankyrase, a poly(ADP-ribose) polymerase at human telomeres. Science (New York, N.Y.), 282(5393), 1484-7. URL http://www.ncbi.nlm.nih.gov/pubmed/9822378.

Steenstrup, T., Hjelmborg, J. V. B., Mortensen, L. H., Kimura, M., Christensen, K., \& Aviv, A. (2013). Leukocyte telomere dynamics in the elderly. European journal of epidemiology, 28(2), 181-7. URL http://www.ncbi.nlm.nih.gov/pubmed/23430034.

Steer, S. E., Williams, F. M. K., Kato, B., Gardner, J. P., Norman, P. J., Hall, M. a., Kimura, M., Vaughan, R., Aviv, A., \& Spector, T. D. (2007). Reduced telomere length in rheumatoid arthritis is independent of disease activity and duration. Annals of the rheumatic diseases, 66(4), 476-80. URL http://www.ncbi.nlm.nih.gov/pubmed/17114192.

Stella, A., Surdo, N. C., Lastella, P., Barana, D., Oliani, C., Tibiletti, M. G., Viel, A., Natale, C., Piepoli, A., Marra, G., \& Guanti, G. (2007). Germline novel MSH2 deletions and a founder MSH2 deletion associated with anticipation effects in HNPCC. Clinical genetics, 71(2), 130-9. URL http: / /www.ncbi.nlm.nih.gov/pubmed/17250661.

Stewart, J. a., Chaiken, M. F., Wang, F., \& Price, C. M. (2012). Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation. Mutation research, 730(1-2), 12-9. URL http://www.ncbi.nlm.nih.gov/pubmed/21945241.

Stindl, R. (2016). The paradox of longer sperm telomeres in older men's testes: a birth-cohort effect caused by transgenerational telomere erosion in the female germline. Molecular cytogenetics, 9(1), 12. URL http://www.ncbi.nlm.nih.gov/pubmed/26858775.

Stupart, D., Goldberg, P., Algar, U., Vorster, A., \& Ramesar, R. (2013). No evidence of genetic anticipation in a large family with Lynch syndrome. Familial cancer. URL http://www.ncbi.nlm.nih.gov/pubmed/23771324.

Stupart, D., Win, A. K., Jenkins, M., Winship, I. M., Goldberg, P., \& Ramesar, R. (2014). Fertility and apparent genetic anticipation in Lynch syndrome. Familial cancer. URL http://www.ncbi.nlm.nih.gov/pubmed/24677027.

Su, S. S. \& Modrich, P. (1986). Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. Proceedings of the National Academy of Sciences of the United States of America, 83(14), 5057-61. URL http://www.ncbi.nlm.nih.gov/pubmed/3014530.

Sun, F. \& Reichenberger, E. J. (2014). Saliva as a source of genomic DNA for genetic studies: review of current methods and applications. Oral health and dental management, 13(2), 217-22. URL http://www.ncbi.nlm.nih.gov/pubmed/24984625.

Supek, F. \& Lehner, B. (2015). Differential DNA mismatch repair underlies mutation rate variation across the human genome. Nature, 521(7550), 81-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/25707793.

Suram, A. \& Herbig, U. (2014). The replicometer is broken: telomeres activate cellular senescence in response to genotoxic stresses. Aging cell, 13(5), 780-6. URL http: //www.ncbi.nlm.nih.gov/pubmed/25040628.

Svenson, U., Nordfjäll, K., Baird, D., Roger, L., Osterman, P., Hellenius, M.-L., \& Roos, G. (2011). Blood cell telomere length is a dynamic feature. PloS one, 6(6), e21485. URL http://www.ncbi.nlm.nih.gov/pubmed/21720548.

Szankasi, P. \& Smith, G. R. (1995). A role for exonuclease I from S. pombe in mutation avoidance and mismatch correction. Science (New York, N.Y.), 267(5201), 1166-9. URL http://www.ncbi.nlm.nih.gov/pubmed/7855597.

Tabori, U., Nanda, S., Druker, H., Lees, J., \& Malkin, D. (2007). Younger age of cancer initiation is associated with shorter telomere length in Li-Fraumeni syndrome. Cancer research, 67(4), 1415-8. URL http://www.ncbi.nlm.nih.gov/pubmed/17308077.

Takagi, S., Kinouchi, Y., Hiwatashi, N., Nagashima, F., Chida, M., Takahashi, S., Negoro, K., Shimosegawa, T., \& Toyota, T. (2000). Relationship between microsatellite instability and telomere shortening in colorectal cancer. Diseases of the colon and rectum, 43(10 Suppl), S12-7. URL http://www.ncbi.nlm.nih.gov/pubmed/11052472.

Takubo, K., Aida, J., Izumiyama, N., Ishikawa, N., Fujiwara, M., Poon, S. S. S., Kondo, H., Kammori, M., Matsuura, M., Sawabe, M., Arai, T., Baird, D. M., \& Nakamura, K.-I. (2010). Chromosomal instability and telomere lengths of each chromosomal arm measured by Q-FISH in human fibroblast strains prior to replicative senescence. Mechanisms of ageing and development, 131(10), 614-24. URL http: //www.ncbi.nlm.nih.gov/pubmed/20817007.

Takubo, K., Izumiyama-Shimomura, N., Honma, N., Sawabe, M., Arai, T., Kato, M., Oshimura, M., \& Nakamura, K.-I. (2002). Telomere lengths are characteristic in each human individual. Experimental gerontology, 37(4), 523-31. URL http: //www.ncbi.nlm.nih.gov/pubmed/11830355.

Theall, K. P., Brett, Z. H., Shirtcliff, E. A., Dunn, E. C., \& Drury, S. S. (2013). Neighborhood disorder and telomeres: connecting children's exposure to community level stress and cellular response. Social science \& medicine (1982), 85(March), 50-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/23540366.

Thiede, C., Prange-Krex, G., Freiberg-Richter, J., Bornhäuser, M., \& Ehninger, G. (2000). Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplants. Bone marrow transplantation, 25(5), 575-7. URL http: / /www.ncbi.nlm.nih.gov/pu.bmed/10713640.

Timshel, S., Therkildsen, C., Bendahl, P.-O., Bernstein, I., \& Nilbert, M. (2009). An effect from anticipation also in hereditary nonpolyposis colorectal cancer families without identified mutations. Cancer epidemiology, 33(3-4), 231-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/19748845.

Tiraby, J. G. \& Fox, M. S. (1973). Marker discrimination in transformation and mutation of pneumococcus. Proceedings of the National Academy of Sciences of the United States of America, 70(12), 3541-5. URL http://www.ncbi.nlm.nih.gov/pubmed/4148702.

Tishkoff, D. X., Boerger, a. L., Bertrand, P., Filosi, N., Gaida, G. M., Kane, M. F., \& Kolodner, R. D. (1997a). Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. Proceedings of the National Academy of Sciences of the United States of America, 94(14), 7487-92. URL http://www.ncbi.nlm.nih.gov/pubmed/9207118.

Tishkoff, D. X., Filosi, N., Gaida, G. M., \& Kolodner, R. D. (1997b). A novel mutation avoidance mechanism dependent on S . cerevisiae RAD27 is distinct from DNA mismatch repair. Cell, 88(2), 253-63. URL http://www.ncbi.nlm.nih.gov/pubmed/9008166.

Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G., \& Vogelstein, B. (2015). Only three driver gene mutations are required for the development of lung and colorectal cancers. Proceedings of the National Academy of Sciences of the United States of America, 112(1), 118-23. URL http://www.ncbi.nlm.nih.gov/pubmed/25535351.

Tong, A. S., Stern, J. L., Sfeir, A., Kartawinata, M., de Lange, T., Zhu, X.-D., \& Bryan, T. M. (2015). ATM and ATR Signaling Regulate the Recruitment of Human Telomerase to Telomeres. Cell reports, 13(8), 1633-46. URL http: //www.ncbi.nlm.nih.gov/pubmed/26586433.

Tsai, Y. Y., Petersen, G. M., Booker, S. V., Bacon, J. a., Hamilton, S. R., \& Giardiello, F. M. (1997). Evidence against genetic anticipation in familial colorectal cancer. Genetic epidemiology, 14(4), 435-46. URL http: //www.ncbi.nlm.nih.gov/pubmed/9271715.

Turner, S. \& Hartshorne, G. M. (2013). Telomere lengths in human pronuclei, oocytes and spermatozoa. Molecular human reproduction, 19(8), 510-8. URL http: //www.ncbi.nlm.nih.gov/pubmed/23519357.

Umar, a., Boyer, J. C., Thomas, D. C., Nguyen, D. C., Risinger, J. I., Boyd, J., Ionov, Y., Perucho, M., \& Kunkel, T. a. (1994). Defective mismatch repair in extracts of colorectal and endometrial cancer cell lines exhibiting microsatellite instability. The Journal of biological chemistry, 269(20), 14367-70. URL http://www.ncbi.nlm.nih.gov/pubmed/8182040.

Umar, A., Koi, M., Risinger, J. I., Glaab, W. E., Tindall, K. R., Kolodner, R. D., Boland, C. R., Barrett, J. C., \& Kunkel, T. A. (1997). Correction of hypermutability, N-methyl-N'-nitro-N-nitrosoguanidine resistance, and defective DNA mismatch repair by introducing chromosome 2 into human tumor cells with mutations in MSH2 and MSH6. Cancer research, 57(18), 3949-55. URL http://www.ncbi.nlm.nih.gov/pubmed/9307278.

Umar, a. \& Kunkel, T. a. (1996). DNA-replication fidelity, mismatch repair and genome instability in cancer cells. European journal of biochemistry / FEBS, 238(2), 297-307. URL http://www.ncbi.nlm.nih.gov/pubmed/8681938.

Unryn, B. M., Cook, L. S., \& Riabowol, K. T. (2005). Paternal age is positively linked to telomere length of children. Aging cell, 4(2), 97-101. URL http: //www.ncbi.nlm.nih.gov/pubmed/15771613.

Valdes, A. M., Andrew, T., Gardner, J. P., Kimura, M., Oelsner, E., Cherkas, L. F., Aviv, A., \& Spector, T. D. (2005). Obesity, cigarette smoking, and telomere length in women. Lancet (London, England), 366(9486), 662-4. URL http: //www.ncbi.nlm.nih.gov/pubmed/16112303.

Valls-Bautista, C., Piñol-Felis, C., Reñé-Espinet, J. M., Buenestado-García, J., \& ViñasSalas, J. (2015). In colon cancer, normal colon tissue and blood cells have altered telomere lengths. Journal of surgical oncology, 111(7), 899-904. URL http: //www.ncbi.nlm.nih.gov/pubmed/25873347.
van den Broek, W. J., Nelen, M. R., Wansink, D. G., Coerwinkel, M. M., te Riele, H., Groenen, P. J., \& Wieringa, B. (2002). Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. Hum Mol Genet, 11(2), 191-198. URL http: //www.ncbi.nlm.nih.gov/pubmed/11809728.
van Steensel, B., Smogorzewska, a., \& de Lange, T. (1998). TRF2 protects human telomeres from end-to-end fusions. Cell, 92(3), 401-13. URL http: //www.ncbi.nlm.nih.gov/pubmed/9476899.

Vasa-Nicotera, M., Brouilette, S., Mangino, M., Thompson, J. R., Braund, P., Clemitson, J.-R., Mason, A., Bodycote, C. L., Raleigh, S. M., Louis, E., \& Samani, N. J. (2005). Mapping of a major locus that determines telomere length in humans. American journal of human genetics, 76(1), 147-51. URL http://www.ncbi.nlm.nih.gov/pubmed/15520935.

Vasen, H. F. a., Möslein, G., Alonso, A., Bernstein, I., Bertario, L., Blanco, I., Burn, J., Capella, G., Engel, C., Frayling, I., Friedl, W., Hes, F. J., Hodgson, S., Mecklin, J.-P., Møller, P., Nagengast, F., Parc, Y., Renkonen-Sinisalo, L., Sampson, J. R., Stormorken, A., \& Wijnen, J. (2007). Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). Journal of medical genetics, 44(6), 353-62. URL http://www.ncbi.nlm.nih.gov/pubmed/17327285.

Vaziri, H., Schächter, F., Uchida, I., Wei, L., Zhu, X., Effros, R., Cohen, D., \& Harley, C. B. (1993). Loss of telomeric DNA during aging of normal and trisomy 21 human lymphocytes. American journal of human genetics, 52(4), 661-7. URL http: //www.ncbi.nlm.nih.gov/pubmed/8460632.

Vera, E., Bernardes de Jesus, B., Foronda, M., Flores, J. M., \& Blasco, M. A. (2012). The rate of increase of short telomeres predicts longevity in mammals. Cell reports, 2(4), 732-7. URL http://www.ncbi.nlm.nih.gov/pubmed/23022483.

Verdun, R. E. \& Karlseder, J. (2007). Replication and protection of telomeres. Nature, 447(7147), 924-31. URL http://www.ncbi.nlm.nih.gov/pubmed/17581575.

Verhulst, S., Aviv, A., Benetos, A., Berenson, G. S., \& Kark, J. D. (2013). Do leukocyte telomere length dynamics depend on baseline telomere length? An analysis that corrects for 'regression to the mean'. European journal of epidemiology, 28(11), 859-66. URL http://www.ncbi.nlm.nih.gov/pubmed/23990212.

Vidaurreta, M., Maestro, M.-L., Rafael, S., Veganzones, S., Sanz-Casla, M.-T., Cerdán, J., \& Arroyo, M. (2007). Telomerase activity in colorectal cancer, prognostic factor and implications in the microsatellite instability pathway. World journal of gastroenterology : WJG, 13(28), 3868-72. URL http://www.ncbi.nlm.nih.gov/pubmed/17657844.
von Zglinicki, T. (2002). Oxidative stress shortens telomeres. Trends in biochemical sciences, 27(7), 339-44. URL http://www.ncbi.nlm.nih.gov/pubmed/12114022.
von Zglinicki, T., Pilger, R., \& Sitte, N. (2000a). Accumulation of single-strand breaks is the major cause of telomere shortening in human fibroblasts. Free radical biology \& medicine, 28(1), 64-74. URL http://www.ncbi.nlm.nih.gov/pubmed/10656292.
von Zglinicki, T., Saretzki, G., Döcke, W., \& Lotze, C. (1995). Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: a model for senescence? Experimental cell research, 220(1), 186-93. URL http://www.ncbi.nlm.nih.gov/pubmed/7664835.
von Zglinicki, T., Serra, V., Lorenz, M., Saretzki, G., Lenzen-Grossimlighaus, R., Gessner, R., Risch, A., \& Steinhagen-Thiessen, E. (2000b). Short telomeres in patients with vascular dementia: an indicator of low antioxidative capacity and a possible risk factor? Laboratory investigation; a journal of technical methods and pathology, 80(11), 1739-47. URL http: //www.ncbi.nlm.nih.gov/pubmed/11092534.

Voskuil, D. W., Vasen, H. F., Kampman, E., \& van't Veer, P. (1997). Colorectal cancer risk in HNPCC families: development during lifetime and in successive generations. National Collaborative Group on HNPCC. International journal of cancer, 72(2), 205-9. URL http : //www.ncbi.nlm.nih.gov/pubmed/9219821.

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(5), 447-9. URL http://www.ncbi.nlm.nih.gov/pubmed/15098033.

Wagner, A. K., Soumerai, S. B., Zhang, F., \& Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. Journal of clinical pharmacy and therapeutics, 27(4), 299-309. URL http: //www.ncbi.nlm.nih.gov/pubmed/12174032.

Wang, F., Pan, X., Kalmbach, K., Seth-Smith, M. L., Ye, X., Antumes, D. M. F., Yin, Y., Liu, L., Keefe, D. L., \& Weissman, S. M. (2013). Robust measurement of telomere length in single cells. Proceedings of the National Academy of Sciences of the United States of America, 110(21), E1906-12. URL http://www.ncbi.nlm.nih.gov/pubmed/23661059.

Wang, Y. \& Qin, J. (2003). MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation. Proceedings of the National Academy of Sciences of the United States of America, 100(26), 15387-92. URL http: //www.ncbi.nlm.nih.gov/pubmed/14657349.

Wang, Z., Rhee, D. B., Lu, J., Bohr, C. T., Zhou, F., Vallabhaneni, H., de SouzaPinto, N. C., \& Liu, Y. (2010). Characterization of oxidative guanine damage and repair in mammalian telomeres. PLoS genetics, 6(5), e1000951. URL http: //www.ncbi.nlm.nih.gov/pubmed/20485567.

Watanabe, Y., Haugen-Strano, A., Umar, A., Yamada, K., Hemmi, H., Kikuchi, Y., Takano, S., Shibata, Y., Barrett, J. C., Kunkel, T. a., \& Koi, M. (2000). Complementation of an hMSH2 defect in human colorectal carcinoma cells by human chromosome 2 transfer. Molecular carcinogenesis, 29(1), 37-49. URL http: //www.ncbi.nlm.nih.gov/pubmed/11020245.

Wei, W., Hemmer, R. M., \& Sedivy, J. M. (2001). Role of p14(ARF) in replicative and induced senescence of human fibroblasts. Molecular and cellular biology, 21(20), 6748-57. URL http://www.ncbi.nlm.nih.gov/pubmed/11564860.

Welsh, K. M., Lu, a. L., Clark, S., \& Modrich, P. (1987). Isolation and characterization of the Escherichia coli mutH gene product. The Journal of biological chemistry, 262(32), 15624-9. URL http://www.ncbi.nlm.nih.gov/pubmed/2824465.

Westphalen, A. a., Russell, A. M., Buser, M., Berthod, C. R., Hutter, P., Plasilova, M., Mueller, H., \& Heinimann, K. (2005). Evidence for genetic anticipation in hereditary non-polyposis colorectal cancer. Human genetics, 116(6), 461-5. URL http: //www.ncbi.nlm.nih.gov/pubmed/15772852.

Wheeler, V. C., Lebel, L.-A., Vrbanac, V., Teed, A., te Riele, H., \& MacDonald, M. E. (2003). Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. Human molecular genetics, 12(3), 273-81. URL http: //www.ncbi.nlm.nih.gov/pubmed/12554681.

Williamson, M. S., Game, J. C., \& Fogel, S. (1985). Meiotic gene conversion mutants in Saccharomyces cerevisiae. I. Isolation and characterization of pms1-1 and pms1-2. Genetics, 110(4), 609-46. URL http://www.ncbi.nlm.nih.gov/pubmed/1202584.

Witty, J. P., McDonnell, S., Newell, K. J., Cannon, P., Navre, M., Tressler, R. J., \& Matrisian, L. M. (1994). Modulation of matrilysin levels in colon carcinoma cell lines affects tumorigenicity in vivo. Cancer research, 54(17), 4805-12. URL http: //www.ncbi.nlm.nih.gov/pubmed/8062282.

Wojdyla, L., Stone, A. L., Sethakorn, N., Uppada, S. B., Devito, J. T., Bissonnette, M., \& Puri, N. (2014). T-oligo as an anticancer agent in colorectal cancer. Biochemical and biophysical research communications. URL http://www.ncbi.nlm.nih.gov/pubmed/24632202.

Wright, D. L., Jones, E. L., Mayer, J. F., Oehninger, S., Gibbons, W. E., \& Lanzendorf, S. E. (2001). Characterization of telomerase activity in the human oocyte and preimplantation embryo. Molecular human reproduction, 7(10), 947-55. URL http: //www.ncbi.nlm.nih.gov/pubmed/11574663.

Wright, W. E., Piatyszek, M. A., Rainey, W. E., Byrd, W., \& Shay, J. W. (1996). Telomerase activity in human germline and embryonic tissues and cells. Developmental genetics, 18(2), 173-9. URL http://www.ncbi.nlm.nih.gov/pubmed/8934879.

Wu, Y., Berends, M. J., Sijmons, R. H., Mensink, R. G., Verlind, E., Kooi, K. a., van der Sluis, T., Kempinga, C., van dDer Zee, a. G., Hollema, H., Buys, C. H., Kleibeuker, J. H., \& Hofstra, R. M. (2001). A role for MLH3 in hereditary nonpolyposis colorectal cancer. Nature genetics, 29(2), 137-8. URL http://www.ncbi.nlm.nih.gov/pubmed/11586295.

Xin, H., Liu, D., \& Songyang, Z. (2008). The telosome/shelterin complex and its functions. Genome biology, 9(9), 232. URL http://www.ncbi.nlm.nih.gov/pubmed/18828880.

Xu, Z., Fallet, E., Paoletti, C., Fehrmann, S., Charvin, G., \& Teixeira, M. T. (2015). Two routes to senescence revealed by real-time analysis of telomerasenegative single lineages. Nature communications, 6(May), 7680. URL http: //www.ncbi.nlm.nih.gov/pubmed/26158780.

Yamada-Fukunaga, T., Yamada, M., Hamatani, T., Chikazawa, N., Ogawa, S., Akutsu, H., Miura, T., Miyado, K., Tarín, J. J., Kuji, N., Umezawa, A., \& Yoshimura, Y. (2013). Ageassociated telomere shortening in mouse oocytes. Reproductive biology and endocrinology : RB\&E, 11, 108. URL http://www.ncbi.nlm.nih.gov/pubmed/24261933.

Yamamoto, H., Adachi, Y., Taniguchi, H., Kunimoto, H., Nosho, K., Suzuki, H., \& Shinomura, Y. (2012). Interrelationship between microsatellite instability and microRNA in gastrointestinal cancer. World journal of gastroenterology, 18(22), 2745-55. URL http://www.ncbi.nlm.nih.gov/pubmed/22719182.

Yan, S., Sorrell, M., \& Berman, Z. (2014). Functional interplay between ATM/ATRmediated DNA damage response and DNA repair pathways in oxidative stress. Cellular and molecular life sciences : CMLS, 71(20), 3951-67. URL http: //www.ncbi.nlm.nih.gov/pubmed/24947324.

Yang, J., Xu, Z.-p., Huang, Y., Hamrick, H. E., Duerksen-Hughes, P. J., \& Yu, Y.-n. (2004). ATM and ATR: sensing DNA damage. World journal of gastroenterology : WJG, 10(2), 155-60. URL http://www.ncbi.nlm.nih.gov/pubmed/14716813.

Yang, W. (2000). Structure and function of mismatch repair proteins. Mutation research, 460(3-4), 245-56. URL http://www.ncbi.nlm.nih.gov/pubmed/10946232.

Zahler, A. M., Williamson, J. R., Cech, T. R., \& Prescott, D. M. (1991). Inhibition of telomerase by G-quartet DNA structures. Nature, 350(6320), 718-20. URL http: //www.ncbi.nlm.nih.gov/pubmed/2023635.

Zavodna, M., Bagshaw, A., Brauning, R., \& Gemmell, N. J. (2014). The accuracy, feasibility and challenges of sequencing short tandem repeats using nextgeneration sequencing platforms. PloS one, 9(12), e113862. URL http: //www.ncbi.nlm.nih.gov/pubmed/25436869.

Zeman, M. K. \& Cimprich, K. a. (2014). Causes and consequences of replication stress. Nature cell biology, 16(1), 2-9. URL http://www.ncbi.nlm.nih.gov/pubmed/24366029.

Zhang, R., Chen, W., \& Adams, P. D. (2007). Molecular dissection of formation of senescenceassociated heterochromatin foci. Molecular and cellular biology, 27(6), 2343-58. URL http://www.ncbi.nlm.nih.gov/pubmed/17242207.

Zhang, X., Mar, V., Zhou, W., Harrington, L., \& Robinson, M. O. (1999). Telomere shortening and apoptosis in telomerase-inhibited human tumor cells. Genes \& development, 13(18), 2388-99. URL http://www.ncbi.nlm.nih.gov/pubmed/10500096.

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(11), 4834-43. URL http://www.ncbi.nlm.nih.gov/pubmed/14690602.

Zhu, X., Han, W., Xue, W., Zou, Y., Xie, C., Du, J., \& Jin, G. (2016). The association between telomere length and cancer risk in population studies. Scientific reports, 6,22243 . URL http://www.ncbi.nlm.nih.gov/pubmed/26915412.

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(6), 1489-98. URL http://www.ncbi.nlm.nih.gov/pubmed/14690602.

Zlatanou, A., Despras, E., Braz-Petta, T., Boubakour-Azzouz, I., Pouvelle, C., Stewart, G. S., Nakajima, S., Yasui, A., Ishchenko, A. a., \& Kannouche, P. L. (2011). The hMsh2hMsh6 complex acts in concert with monoubiquitinated PCNA and Pol $\eta$ in response to oxidative DNA damage in human cells. Molecular cell, 43(4), 649-62. URL http: //www.ncbi.nlm.nih.gov/pubmed/21855803.


[^0]:    
    
    

    FISH: fluorescent In Situ Hybridisation• PBL. Peripheral Blood Leukocytes; MZ: Monozygotic Twins; DZ: Dizygotic Twins; PBMC
    

[^1]:    Table 6.4: Age-adjusted telomere length ( $\Delta \mathrm{Tel}$ ) and parent/children pair ( $\mathbf{P} / \mathbf{C}$ ) comparisons for control families.

[^2]:    Red colour indicates $\mathrm{P} / \mathrm{C}$ comparisons where children have shorter are-adjusted telomere length than their parents. Table code: Genes: $1=M L H 1 ; 2=M S H 2 ; 0=$ wild type. Sex: $\mathrm{F}=$ female $\mathrm{M}=$ male. Cancer: CRC=colorectal cancer; MM=malignant melanoma; EC=endometrial cancer; SCC=squamous cell carcinoma; BCC=basal cell carcinoma; UBC=urinary bladder cancer; $\mathrm{TC}=$ thyroid cancer. Age $\mathrm{C}=$ age at first cancer onset. $\mathrm{N} / \mathrm{A}=$ not applicable.

