

INVESTIGATING THE IMPACT OF TELOMERE DYSFUNCTION ON THE CHRONIC LYMPHOCYTIC LEUKAEMIA GENOME

Laura Escudero-Monreal

A thesis submitted for the degree of Doctor of Philosophy

September 2017



Institute of Cancer and Genetics

School of Medicine

Cardiff University

Funded by Cancer Research Wales and Cardiff University

I dedicate this Ph.D. thesis to everyone who suffers from cancer,
for I hope one day, science can fight and win this battle.

Laura Escudero-Monreal

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors Professor Duncan Baird and Dr Kate Liddiard for the opportunity to undertake this PhD. I am very grateful for your support and guidance throughout the project. I would also like to thank Professor Chris Pepper for your advice, to the CLL patients and to Professor Chris Fegan for providing the patient samples.

Thank you to Dr Kate Liddiard, Dr Rhiannon Robinson and Mrs Julia Grimstead for teaching me the techniques and for always being willing to help; to Dr Kevin Norris for facilitating the identification of the CLL patient samples with short telomeres; and to Dr Kez Cleal for your bioinformatics input. To everyone in the lab, I will always remember the lunch time laughs and the fun nights out. Particularly the events to raise money for charity including the CGB Bake Off, the 5K pretty muddy races with The GenetChicks and the Christmas secret Santa. It has been the best working environment and I will miss you all. Especially Sam Hyatt and Melanie Hurtz, thank you for being by my side during this journey; as well as my colleagues and my housemates, you became my best friends.

Sobre todo, quiero dar las gracias a mi familia, por apoyar mis decisiones y ayudarme durante el camino a convertirme en investigadora. Sin vuestro cariño, ayuda y entendimiento, hoy no estaría aquí. A mis amigas más cercanas, gracias por recordarme que hay vida más allá de la academia y ayudarme a mantener un equilibrio en mi vida. Junto con mi familia, gracias especialmente por darme ánimos cuando más los necesitaba. To my partner Jahn, thank you for your understanding, for supporting my decisions and for your unconditional love. Thank you to your family, for always making me feel welcome.

Thank you to Cancer Research Wales (CRW) and Cardiff University for the funding that has enabled this project. I am grateful to the European Association for Cancer Research (EACR), Cancer Research UK (CRUK), Keystone Symposia and The Genetics Society for the travel awards that allowed me to present my work at several conferences. It helped me to broaden my horizons, get new ideas and establish collaborations.

To all of you, thank you. During these past 4 years I have grown as a person but especially as a scientist. Now, more than ever, I want to become an excellent researcher and help in the fight against cancer. I know it will be hard, but with passion and dedication, everything is possible.

SUMMARY

Short dysfunctional telomeres can result in chromosome fusions that drive genomic rearrangements and ultimately malignant progression. In Chronic Lymphocytic Leukaemia (CLL), telomere length (TL) is a powerful predictor of patient survival. The aim of this project is to understand the role that telomere dysfunction plays in driving the evolution of the CLL genome.

Telomere fusions were detected in 71% of 276 CLL patients with short telomeres (TL<3.81Kb). From 9 CLL patient samples with the highest fusion frequency ($>4.20 \times 10^{-5}$ /diploid genome), 914 telomere fusions were characterised using Illumina HiSeq paired-end sequencing. In addition to intra- and inter-chromosomal recombinations, telomere fusions with non-telomeric loci were detected, including the ancestral telomere at Chr2q13, mitochondrial DNA, and loci associated with copy number aberrations in CLL. Telomere fusions also incorporated genes expressed in CLL-B lymphocytes and other oncogenes, suggesting that active chromatin is more prone to damage and aberrant repair. These events were potentially mediated by A-NHEJ that requires microhomology.

Translocations involving *hTERT*, proximal to the 5p telomere, have previously been detected in CLL and associated with telomerase upregulation. In this study, 5p telomere fusions were identified in 22.6% patient samples and 172 fusion events that involved 5p were characterised, which may explain mechanisms of telomerase reactivation in cancer.

Surprisingly, 67% of the 9 patients presented bimodal TL distributions compared to the overall cohort (4% of 276), consistent with intra-tumour heterogeneity, which was confirmed for one patient. For this patient with indolent disease, chromosomal rearrangements were detected, in addition to a novel mutation in *REV3L* implicated in translesion synthesis that may negatively impact cancer cells fitness and be a potential therapeutic target.

This study shows that telomere dysfunction in CLL initiates genome-wide instability providing a source for genetic variability that allows intra-tumour heterogeneity and tumour progression. However, therapeutically-exploiting this instability could prove beneficial for patient outcome.

LIST OF ABBREVIATIONS

ADMIRE	Does the ADdition of Mitoxantrone Improve Response to FCR chemotherapy in patients with CLL? (ADMIRE) trial
AID	APOBEC-Induced cytosine Deamination (c: canonical; nc: non-canonical)
ALC	Absolute Lymphocyte Count
ALT	Alternative Lengthening of Telomeres
AML	Acute Myeloid Leukaemia
ARCTIC	Attenuated dose Rituximab with ChemoTherapy In Chronic lymphocytic leukaemia (ARCTIC) trial
A-NHEJ	Alternative Non-Homologous End Joining
ANOVA	ANalysis OF VAriance
APBs	ALT-associated promyelocytic leukaemia (PML) nuclear bodies
ATM	Ataxia Telangiectasia Mutated
ATR	Ataxia telangiectasia and Rad3 related
ATRX	Alpha Thalassemia/Mental Retardation Syndrome X-Linked
BAI	Indexed BAM file
BAM	Binary version of a SAM file
BFB	Breakage-Fusion-Bridge
BIR	Break-Induced Replication
BL	Baseline
BLAST	Basic Local Alignment Search Tool
BML	Bloom syndrome protein
bp	Base Pairs
BRCA1	Breast Cancer Type 1 Susceptibility
BRCA2	Breast Cancer Type 2 Susceptibility
BSA	Bovine Serum Albumin
BTK	Bruton Tyrosine Kinase
BWA	Burrows-Wheeler Aligner
CBS	Central Biotechnology Service (CBS)
CDK	Cyclin-Dependent Kinase
CF	Chromosome Fusion
CFS	Common Fragile Site
CHK1	Checkpoint Kinase 1
CHK2	Checkpoint Kinase 2
CIN	Chromosomal INstability
CLL	Chronic Lymphocytic Leukaemia
C-NHEJ	Classical Non-Homologous End Joining
CN-LOH	Copy-Neutral Loss Of Heterozygosity
CNV	Copy Number Variation
CTCF	CCCTC-binding factor
CtIP	CtBP-interacting protein
DAPI	4',6-diamidino-2-phenylindole

DAVID	Database for Annotation, Visualization and Integrated Discovery
DAXX	Death-Domain Associated
DDR	DNA-Damage Response
DKC	Dyskerin
DLBCL	Diffuse Large B-Cell Lymphoma
D-Loop	Displacement Loop
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-Dependent Protein Kinase
DN- <i>hTERT</i>	Dominant Negative Human Telomerase Reverse Transcriptase
dNTP	Deoxyribonucleotide
DSB	Double-Strand Break
ECTR	Extrachromosomal telomeric repeats
EDTA	Ethylenediaminetetraacetic Acid
EtBr	Ethidium Bromide
ETS	E-twenty-six-domain
EXO1	Exonuclease 1
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FCS	Fetal Calf Serum
FCR	Fluradabine, Chlorabucil and Rituximab
FISH	Fluorescence <i>In situ</i> Hybridisation
FJ	Fusion Junction (m: mapped; um: unmapped)
FoSTeS	Fork Stalling and Template Switching
GABP	GA-Binding Protein
GRCh37	Genome Reference Consortium human build 37
GSEA	Gene Set Enrichment Analysis
GWAS	Genome-Wide Association Studies
H2AFX	H2A histone Family, member X
HCl	Hydrochloric Acid
HCT1080	Human fibrosarcoma cell line
HCT116	Human colon cancer cell line
HDR	Homology Directed Repair
HEK293	Human Embryonic Kidney 293 cell line
hg19	Human genome version 19
HJ	Holliday Junction
hnRNPA1	heterogeneous nuclear RiboNucleoProtein A1
HR	Homologous Recombination
HR	Hazard Ratio
HSCT	Haematopoietic Stem Cell Transplantation
HT Q-FISH	High Throughput Quantitative Fluorescent <i>In situ</i> Hybridisation
ICGC	International Cancer Genome Consortium
Ig	Immunoglobulin
<i>IGHV</i>	ImmunoGlobulin Heavy-chain Variable region gene (m:mutated; u: unmutated)
IGV	Integrative Genome Viewer

kb	Kilobases
Ku	Ku70-Ku80 heterodimer (Ku)
LDT	Lymphocyte Doubling Time
LIG3	DNA ligase 3
LIG3 ^{-/-} -mL3+nuc. LIG3	Depleted nuclear LIG3 and complemented with mitochondrial isoform, and LIG3 complemented to rescue the phenotype
LIG4	DNA ligase 4
LOH	Loss Of Heterozygosity
M1	Mortality stage 1
LRF CLL4	Leukaemia Research Fund CLL4 trial
M2	Mortality stage 2
MACS	Magnetic-activated cell sorting
MAPK	Mitogen-Activated Protein Kinase
MAPK/ERK	RAS-RAF-MEK-ERK signalling pathway
MAPQ	MAPping Quality
MBL	Monoclonal B cell Lymphocytosis
MDR	Minimal Deleted Region
MDS	MyeloDysplastic Syndrome
MEC1	Human B-CLL cell line
MH	MicroHomology
MM	Multiple Myeloma
MMBIR	Microhomology-Mediated BIR
MMEJ	Microhomology-Mediated End Joining
MMQ-PCR	Monochrome Multiplex Quantitative Polymerase Chain Reaction
MRC5	Human lung fibroblast cell line
MRN	MRE11-RAD50-NBS1
MSigDB	Molecular Signatures Database
MTS	MultiTelomeric Signals
MW	Molecular Weight
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NHP2	Non-Histone Protein 2
NOP10	Nucleolar protein 10
NUMT	Nuclear mitochondrial (mt) DNA
OB	Oligonucleotide/oligosaccharide-Binding
OH	Hydroxyl group
OR	Olfactory Receptors
OS	Overall Survival
PARP	Poly (ADP-Ribose) Polymerase
PBMC	Peripheral Blood Mononuclear Cells
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PD	Population Doubling

PDX	Patient-Derived Xenograft
PE	Paired-End
PFS	Progression-Free Survival
PI3K	Phosphoinositide 3-Kinase
PK	PseudoKnot-template core
Plk3	Polo-like kinase 3
PML	ProMyelocytic Leukaemia
Pol	Polymerase
POT1	Protection Of Telomeres 1
PROVEAN	PROtein Variation Effect ANalyzer
QC	Quality Control
Q-FISH	Quantitative Fluorescent <i>In situ</i> Hybridisation
Q-PCR	Quantitative Polymerase Chain Reaction
RAP1	Repressor Activator Protein 1
Rb	Retinoblastoma
R-loop	TERRA RNA-telomeric DNA hybrid
RNA	Ribonucleic Acid
RNaseH2	Ribonuclease H2
ROS	Reactive Oxygen Species
RPA	Replication Protein A
RT	Reverse Transcriptase
RTEL1	Regulator of Telomere Elongation helicase 1
SAM	Sequence Alignment Map
SC	Sister Chromatid fusions
SCE	Sister Chromatid Exchange
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SE	Standard Error
SHM	Somatic HyperMutation
SIFT	Sorting Tolerant from Intolerant
SNP	Single-Nucleotide Polymorphism
SNV	Single-Nucleotide Variation
SSA	Single-Strand Annealing
STE	Stem-loop
STELA	Single TELomere Length Analysis
TAE	Tris-Acetate-EDTA
TCAB1	Telomerase CAjal Body protein 1
TCF	Ternary Complex Factor
T-circles	Telomere circles
TEL patch	TPP1 glutamate (E) and leucine (L)-rich patch
TEN	Telomerase Essential N-Terminal
TERC	TELomerase RNA Component
TERRA	TELomeric Repeat-Containing RNA
TERT	TELomerase Reverse Transcriptase
TFUs	Telomere Fluorescent Units
TIF	Telomere-Induced DNA damage Foci

TIN2	TRF1-Interacting Nuclear Factor 2
TL	Telomere length
T-Loop	Telomere Loop
T _m	Melting temperature
TMM	Telomere Maintenance Mechanism
TNK1	Tankyrase-1
TNK2	Tankyrase-2
<i>TP53</i>	Tumor suppressor <i>P53</i> gene
TPE	Telomeric Position Effect
TPE-OLD	Telomeric Position Effect Over Long Distance
TPM	TERT Promotor Mutations
TPP1	TIN2/PTOP/PIP1
TRB	Telomerase RNA Binding
TRF	Terminal Restriction Fragment
TRF1	Telomeric Repeat-Binding Factor 1
TRF2	Telomeric Repeat-Binding Factor 2
Ts/Tv	Base transitions (Ts) and base transversions (Tv) ratio
TTFT	Time To First Treatment
TVR	Telomeric Variant Repeat
TZAP	Telomeric Zinc finger-Associated Protein
UHW	University Hospital of Wales
UPD	UniParental Disomy
USCS	University of California, Santa Cruz
UV	UltraViolet
VAF	Variant Allele Frequency
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organization
WT	Wild Type
XRCC4	X-ray Repair Cross-Complementing protein 4
53BP1	p53-binding protein 1

LIST OF CONTENTS

NOTICE OF SUBMISSION	III
ACKNOWLEDGEMENTS	IV
SUMMARY	V
LIST OF ABBREVIATIONS	VI
LIST OF CONTENTS	XI
LIST OF FIGURES AND TABLES	XX
CHAPTER 1: INTRODUCTION	1
1.1 The telomere system.....	1
1.1.1 The telomeric sequence.....	2
1.1.2 The Shelterin complex	3
1.1.3 Telomere protection: hiding the end and inhibiting DSB response	3
1.1.4 DNA damage response and DNA repair mechanisms at telomeres	5
1.1.4.1 Classical non-homologous end joining (C-NHEJ)	6
1.1.4.2 Alternative non-homologous end joining (A-NHEJ).....	7
1.1.4.3 Single-strand annealing (SSA)	7
1.1.4.4 Homologous recombination (HR).....	8
1.1.4.5 Choice of pathway and implications in cancer	8
1.1.4.6 DNA repair activity at telomeres	9
1.1.5 Telomere maintenance mechanisms.....	11
1.1.5.1 Telomerase	11
1.1.5.2 Telomere shortening and senescence: a tumour suppressor mechanism.....	12
1.1.5.3 Telomere length regulation by modulating telomerase.....	15
1.1.5.4 Alternative lengthening of telomeres (ALT)	15

1.1.6	Subtelomeric DNA, gene families, TERRA and telomere-position effect	17
1.2	Telomeres and telomerase in cancer	19
1.2.1	Telomere crisis: a source of genomic instability and malignant transformation	19
1.2.2	Reactivation of telomere maintaining mechanisms in cancer	22
1.2.3	Telomere length and telomere dysfunction: a powerful prognostic marker ...	23
1.2.3.1	Terminal Restriction Fragment (TRF) analysis	23
1.2.3.2	Quantitative Polymerase Chain Reaction (Q-PCR)	24
1.2.3.3	Quantitative fluorescence <i>in situ</i> hybridisation (Q-FISH)	24
1.2.3.4	Flow-FISH	25
1.2.3.5	Single TELomere Length Analysis (STELA)	25
1.2.3.6	Detecting telomere fusions	26
1.3	Telomere length in Chronic Lymphocytic Leukaemia	28
1.3.1	Chronic lymphocytic leukaemia	28
1.3.1.1	Symptoms, diagnosis and staging	28
1.3.1.2	Prognostic markers	29
1.3.2	Telomere length is a prognostic marker in CLL: 20 years of knowledge	30
1.4	The CLL cancer genome	33
1.4.1	CLL driver mutations and cytogenetic aberrations	33
1.4.2	Tumour heterogeneity	36
1.4.3	Other 'Omics' studies	37
1.4.4	Therapeutics and precision medicine	37
1.5	Hypothesis and aims of the project	39
	CHAPTER 2: MATERIALS AND METHODS	40
2.1	MATERIALS	40
2.1.1	Chemicals and reagents	40
2.1.2	Laboratory materials and equipment	40

2.1.3	Oligonucleotides	40
2.1.4	CLL patient samples	41
2.2	METHODS	42
2.2.1	Isolation of PBMCs	42
2.2.2	Cell culture	42
2.2.3	Cell count and viability	43
2.2.4	Cryopreservation of cells and thawing procedure	43
2.2.5	Cell enrichment using magnetic microbeads	43
2.2.6	Cell sorting by flow cytometry	44
2.2.7	DNA extraction: Phenol/Chloroform/Isoamyl Alcohol	44
2.2.8	DNA quantification	45
2.2.9	Polymerase Chain Reaction (PCR)	46
2.2.9.1	Conventional PCR	46
2.2.9.2	STELA PCR	47
2.2.9.3	TVR PCR	48
2.2.9.4	Telomere fusion PCR	49
2.2.9.5	Fusion reamplification PCR	50
2.2.10	Gel electrophoresis	51
2.2.10.1	Gel electrophoresis for STELA, TVR and fusion PCR products	51
2.2.10.2	Standard PCR products	51
2.2.10.3	Visualisation of PCR products	51
2.2.11	Southern blotting	52
2.2.11.1	DNA transfer from an agarose gel to a nylon membrane	52
2.2.11.2	Synthesis of the radiolabelled DNA probe	52
2.2.11.3	Hybridisation	52
2.2.11.4	Removing unbound probe	53
2.2.11.5	Visualisation of radiolabelled blots	53
2.2.12	Gel analysis and statistics	54

2.2.13	Sanger Sequencing.....	54
2.2.13.1	DNA extraction from agarose gel	54
2.2.13.2	Sequencing reaction	54
2.2.13.3	Analysis of sequences	55
2.2.14	Paired-end Next Generation Sequencing	55
2.2.14.1	Pre-sequencing PCR clean-up	55
2.2.14.2	Illumina HiSeq4000 PE100 paired-end NGS	56
2.2.14.3	Data processing and quality controls (QC) with bioinformatics tools	57
2.2.14.4	Manual curation and downstream analysis.....	58
2.2.15	Whole Genome Sequencing	60
2.2.15.1	Illumina HiSeq2000.....	60
2.2.15.2	Data processing and quality controls (QC) with bioinformatics tools	60
2.2.15.3	Variant calling: SNVs and copy numbers	60
2.2.15.4	Manual curation and downstream analysis.....	60
CHAPTER 3: CHARACTERISATION OF THE 5P TELOMERE: MEASUREMENT OF THE TELOMERE LENGTH AND INVESTIGATION OF TELOMERE DYSFUNCTION AND FUSION. ...		61
3.1	Abstract.....	61
3.2	Introduction	62
3.3	Aims of the project.....	63
3.4	Results	64
3.4.1	5p STELA measures the telomere length of the 5p chromosome arm in human cell lines and CLL patients.....	64
3.4.2	Comparison of the telomere lengths at 5p, 17p and XpYp in CLL patients	68
3.4.3	The 5p telomere contains more variant repeats than the XpYp telomere	71
3.4.4	5p telomere fusions are detected with the 5p8:17p6:XpYpM:16p1:21q1 telomere fusion assay in human cell lines and CLL patients	73

3.4.4.1	Development of the telomere fusion assay at the 5p chromosome end.....	73
3.4.4.2	Optimisation: testing different telomere fusion primer combinations	76
3.4.4.3	The 5p8 PCR artefact is generated by non-specific binding of the 5p8 primer	80
3.4.4.4	The inclusion of the 5p telomere in the fusion assay increases the detection of telomere fusion amplicons	81
3.4.5	Characterisation of the 5p telomere in a HCT116 ^{DN-hTERT} clone progressing through a telomere-driven crisis.	83
3.4.5.1	Characterisation of the 5p telomere dynamics using STELA	85
3.4.6	Telomere dysfunction and fusion in CLL, including the 5p telomere	90
3.4.6.1	CLL patient samples exhibit different frequencies of telomere fusions, including the 5p telomere	91
3.4.6.2	The 5p telomere undergoes fusion in CLL	93
3.4.6.3	The frequency of fusions may provide prognostic information in CLL patients with short telomeres.....	94
3.5	Discussion.....	98
3.5.1	The 5p telomere length distributions are similar to the XpYp and 17p telomeres.....	98
3.5.2	The Fusion assay at the 5p chromosome end	100
3.5.4	The 5p telomere dynamics on HCT116 ^{DN-hTERT} before, during and after crisis are similar to the XpYp telomere.....	101
3.5.5	Telomere dysfunction and fusion in CLL, including the 5p telomere.	102
3.5.6	Telomere fusion frequency and prognosis in CLL patients with short telomeres.....	103
3.5.7	Summary.....	106

CHAPTER 4: INVESTIGATING THE IMPACT OF TELOMERE DYSFUNCTION ON THE CHRONIC LYMPHOCYTIC LEUKAEMIA GENOME	107
4.1 Abstract	107
4.2 Introduction	108
4.3 Aims of the project.....	110
4.4 Results	111
4.4.1 Specialised paired-end NGS of telomere fusion amplicons.....	111
4.4.1.1 9 CLL patient samples with the highest frequency of fusions selected for specialised paired-end Illumina HiSeq4000 PE100....	111
4.4.1.2 Mapping pipeline: intra- and inter-chromosomal analysis	114
4.4.1.3 Classification of events identified.....	115
4.4.1.4 Validating intra-chromosomal fusion events	116
4.4.1.5 Validating inter-chromosomal fusion events	121
4.4.2 Signature of intra- and inter-chromosomal telomere fusion events in a panel of 9 CLL patient samples	124
4.4.3 Variety of TVR identified from the intra analysis.....	129
4.4.5 Subtelomeric deletion and asymmetry in sister chromatid and intra-chromosomal fusion events in CLL samples.	131
4.4.6 Distinct microhomology utilisation at the fusion junction	133
4.4.7 Investigating telomere fusion activity with non-telomeric genomic loci	136
4.4.7.1 Telomeric fusions with mitochondrial DNA.....	138
4.4.7.2 Fragility at the ancestral telomere at Chr2q13.....	139
4.4.7.3 Characteristics of genomic loci that fuse with dysfunctional telomeres.....	141
4.4.7.4 B lymphocyte specific-genes and other oncogenes are hotspots of telomere fusion activity in CLL	143
4.5 Discussion.....	145
4.5.1 Higher frequency of bimodal telomere length distributions in CLL-B cells with the highest frequency of fusions	145

4.5.2	Signature of intra- and inter-chromosomal telomere fusion events in a panel of 9 CLL patient samples	146
4.5.3	Variety of TVR identified from the intra-chromosomal analysis	147
4.5.4	Asymmetric deletion at sister chromatids in intra-chromosomal fusion events.....	148
4.5.5	Distinct degrees of microhomology at the fusion junction	148
4.5.6	Investigating hotspots of telomere fusion activity with genomic loci.....	150
4.5.6.1	Telomeric fusions with mitochondrial DNA.....	150
4.5.6.2	Fragility at the ancestral telomere at Chr2q13.....	152
4.5.6.3	Characteristics of genomic loci that fuse with dysfunctional telomeres.....	153
4.5.6.4	B lymphocyte specific-genes and other oncogenes are hotspots of telomere fusion activity in CLL	154
4.5.7	Summary.....	156

CHAPTER 5: WHOLE GENOME SEQUENCING IN A PATIENT WITH CHRONIC LYMPHOCYTIC LEUKAEMIA..... 157

5.1	ABSTRACT	157
5.2	Introduction	158
5.3	Aims of the project.....	160
5.4	Results	161
5.4.1	Whole Genome Sequencing of a CLL patient sample with a high frequency of telomere fusions and bimodal TL distribution.	161
5.4.1.1	CLL patient sample with bimodal TL distribution	161
5.4.1.2	Preparation of the CLL patient sample for WGS.....	163
5.4.2	Intra-tumour heterogeneity detected in a CLL patient sample.....	166
5.4.2.1	Bioinformatics sequence analysis: Identifying SNVs	166
5.4.2.2	SNV analysis reveals multiclonality.....	166

5.4.2.3	Validation of tumour multiclonality by STELA and Sanger sequencing.....	168
5.4.3	Mutation in CLL drivers and genes associated with telomere dysfunction....	171
5.4.3.1	Identification of mutations: Validation pipeline.....	171
5.4.3.2	Impact of mutations: CLL drivers and the MAPK/ERK pathway	174
5.4.3.3	Telomeres and POT1 mutations	176
5.4.4	The signature of mutations in DB17 CLL-B cells is characterised by T>G transversions and associated with nc-AID.....	177
5.4.4.1	Transition/Transversion ratio	177
5.4.4.2	Signature of mutations	178
5.4.4.3	Localised hypermutation is exclusive to the Immunoglobulin locus.....	179
5.4.5	Mutated <i>IGHV</i> gene or multiple clones with diverse <i>IGHV</i> status?	182
5.4.6	The common CLL 13q deletion arising from a 6q:13q translocation.....	184
5.4.7	Copy neutral loss of heterozygosity: 17p uniparental disomy (17p UPD) with mutated <i>TP53</i>	186
5.5	Discussion.....	189
5.5.1	Multiclonality: theories.....	189
5.5.2	Impact of mutations	190
5.5.2.1	CLL drivers and the MAPK/ERK signalling pathway	190
5.5.2.2	Truncated REV3L and chromosome instability.....	192
5.5.3	POT1 and telomeres	193
5.5.4	Mutated CLL and a signature of mutations consistent with nc-AID.....	194
5.5.5	6q:13q translocation leads to 13q14 and 6q21 deletion in CLL patient.....	195
5.5.6	17p UPD, <i>TP53</i> and telomeres	196
5.5.7	Excessive genetic instability may negatively impact on cancer cell fitness resulting in patient's indolent disease.....	199

CHAPTER 6: GENERAL DISCUSSION AND FUTURE DIRECTIONS	202
6.1 Summary	202
6.2 General discussion	204
6.2.1 Differences in telomere length at distinct chromosome ends	204
6.2.2 Telomere dysfunction and fusion, including the 5p telomere, in CLL patient samples and in HCT116 human colorectal cancer cell line	205
6.2.3 Telomere fusion frequency and prognosis in CLL patients with short telomeres.....	207
6.2.4 The impact of telomere dysfunction and fusion on the CLL genome.....	208
6.2.5 Telomere dysfunction and fusion potentially drive clonal evolution and intra-tumour heterogeneity.....	209
6.2.6 Excessive genetic instability and decreased tolerance to DNA damage may negatively impact on cancer cell survival, improving patient outcome	210
6.2.7 Potential therapeutic interventions to prevent cells from escaping crisis	211
6.3 Conclusions and future directions	213
6.3.1 Final conclusions	217
APPENDIX	219
Supplementary Figures	219
Supplementary Tables	226
REFERENCES.....	250

LIST OF FIGURES AND TABLES

LIST OF FIGURES

Figure 1.1. The human telomeric system.	1
Figure 1.2. The structure of telomeric DNA.	4
Figure 1.3. Synthesis of DSB and SSB repair mechanisms.	6
Figure 1.4. Telomere length elongation by telomerase.	12
Figure 1.5. The end-replication problem.	13
Figure 1.6. Templates for HR in ALT.	16
Figure 1.7. Breakage-Fusion-Bridge cycle.	20
Figure 1.8. The role of telomeres in ageing and cancer.	21
Figure 1.9. Representation of STELA PCR.	25
Figure 1.10. Telomere fusion PCR.	26
Figure 1.11. Mutated pathways in CLL.	34
Figure 2.1. Paired-end sequencing and alignment.	56
Figure 2.2. Workflow for the NGS analysis of telomere fusion amplicons.	59
Figure 3.1. Development of 5p STELA.	64
Figure 3.2. Validation of the specificity and optimisation of 5p STELA.	65
Figure 3.3. 5p STELA on 12 CLL patient samples.	67
Figure 3.4. Comparison of the telomere lengths at 5p, 17p and XpYp.	69
Figure 3.5. Comparison of the telomere length at 5p, 17p and XpYp for 57 CLL patients. ...	70
Figure 3.6. Comparison of the Telomere Variant Repeat at 5p and XpYp.	72
Figure 3.7. Chromosome 5p sequence and localisation of the primers.	74
Figure 3.8. Primer design at the 5p chromosome end.	75
Figure 3.9. GC distribution at the 5p subtelomere.	75
Figure 3.10. Testing the 5p9 primer into the fusion assay.	77
Figure 3.11. Testing the 5p8 primer into the fusion assay.	78
Figure 3.12. Highlights from the optimisation of the 5p8:17p6:XpYpM:16p1:21q1 fusion assay.	79
Figure 3.13. Characterisation of the 5p8 PCR artefact.	80
Figure 3.14. Comparison of the 17p6:XpYpM:16p1:21q1 with the 5p8:17p6:XpYpM:16p1:21q1 fusion assay for HEK293 cell line at 62°C.	82
Figure 3.15. HCT116 ^{DN-hTERT} cell lines.	84

Figure 3.16. Characterisation of the 5p telomere using STELA.	86
Figure 3.17. Comparison of the 5p to the XpYp telomere.	87
Figure 3.18. Comparison of the 5p to the XpYp molecules amplified using STELA.	88
Figure 3.19. 5p8:17p6:XpYpM:16p1:21q1 fusions on HCT116 DN- <i>hTERT</i> cell lines.	89
Figure 3.20. CLL patient samples exhibit different frequencies of telomere fusions.	92
Figure 3.21. Detection of 5p fusions in CLL patient samples.	93
Figure 3.22. Telomere length, frequency of fusions and ALC.	94
Figure 3.23. Frequency of telomere fusions as a prognostic marker in patients with short telomeres.	96
Figure 3.24. Frequency of fusions and other biomarkers.	97
Figure 4.1. Preparation of 9 CLL patient samples for specialised NGS of telomere fusions	112
Figure 4.2. Bimodality identified in the CLL patient samples for NGS.	113
Figure 4.3. Characterisation of a telomere-genomic inter-chromosomal fusion event.	115
Figure 4.4. Summary of intra-chromosomal fusion analysis.	117
Figure 4.5. Example of the four types of events obtained from the 5p intra-chromosomal fusion analysis.	118
Figure 4.6. Example of the four types of events obtained from the 17p intra-chromosomal fusion analysis.	119
Figure 4.7. Example of the four types of events obtained from the XpYp intra-chromosomal fusion analysis.	120
Figure 4.8. Telomeric inter-chromosomal fusion event.	122
Figure 4.9. Summary of inter-chromosomal fusion analysis.	123
Figure 4.10. Sister chromatid fusion event detected from the inter-chromosomal analysis	123
Figure 4.11. Characterisation of all telomere fusion events detected in 9 CLL patients.	124
Figure 4.12. Characterisation of telomere fusions detected from CLL patient DB17.	125
Figure 4.13. Characterisation of telomere fusions for 8 CLL patients	126
Figure 4.14. Proportion of 5p telomere fusion events.	127
Figure 4.15. Complex telomere fusion events.	128
Figure 4.16. Telomere Variant Repeats identified.	130
Figure 4.17. Asymmetrical deletion prior 5p, 17p or XpYp sister chromatid fusion.	132
Figure 4.18. Distinct degree of microhomology at the fusion junction.	133
Figure 4.19. Long tracts of microhomology at inter-chromosomal fusion junctions with genomic loci.	135

Figure 4.20. Inter-chromosomal telomere-genomic fusions from 9 CLL patient samples. .	137
Figure 4.21. Telomeric fusion with mitochondrial DNA.....	138
Figure 4.22. Telomeric fusions with the ancestral telomere at 2q13.....	140
Figure 4.23. DNA features of genomic locus that fuses with dysfunctional telomeres.	141
Figure 4.24. Telomere fusion with genomic loci 12q13.2.....	144
Figure 5.1. Telomere length at the XpYp, 2p, 12q and 18q.	161
Figure 5.2. XpYp telomere length, 7 year follow up on a CLL patient.	162
Figure 5.3. Isolation of CD19 ⁺ CLL-B cells (tumour) and CD3 ⁺ T cells (control).....	164
Figure 5.4. Comparison of the TL profile for the tumour and control fraction.	165
Figure 5.5. Intra-tumour heterogeneity detected from SNP analysis.	167
Figure 5.6. TL obtained from DB17 CLL cells in culture.	169
Figure 5.7. Validating multiclonality by sequencing SNVs in DB17 at baseline and day 28.	170
Figure 5.8. Summary of SNVs identified from the tumour and control sample.	171
Figure 5.9. Pipeline for validation of SNVs.....	172
Figure 5.10. Significantly mutated genes and pathways in CLL.....	175
Figure 5.11. Point mutations identified in the human POT1 shelterin protein.	176
Figure 5.12. Signature of mutations.	178
Figure 5.13. Proximity of mutation clusters (Kataegis).....	179
Figure 5.14. Kataegis and CNV at 2p11.2.....	180
Figure 5.15. <i>IGHV</i> status in DB17 CLL patient sample.....	183
Figure 5.16. Common 13q deletion originated from 6q:13q tranlocation	185
Figure 5.17. 17p Copy Neutral Loss of Heterozygosity (CN-LOH).	187
Figure 5.18. DB17 CLL patient sample with a homozygous <i>TP53</i> mutation.	188
Figure 5.19. Predicted evolution of the XpYp TL for the patient CLL-B cells.	190
Figure 5.20. REV3L p.Gln1426* mutation causes loss of polymerase domain and REV7 binding site.	192
Figure 5.21. Proposed model explaining the origin of the 17p UPD in the patient CLL-B cells.....	198
Figure 5.22. The paradoxical effect of excessive genetic instability in cancer progression.	201
Figure 6.1. The impact of telomere dysfunction and fusion on Chronic Lymphocytic Leukaemia.	218

LIST OF TABLES

Table 1.1. Binet and Rai classification systems.....	29
Table 1.2. Recurrent copy number changes in CLL.....	35
Table 2.1. Ethical approval for each CLL cohort.....	41
Table 2.2. Conventional PCR reaction mix.....	46
Table 2.3. Conventional PCR cycling conditions.....	46
Table 2.4. STELA PCR reaction mix.....	47
Table 2.5. STELA PCR cycling conditions.....	47
Table 2.6. TVR PCR reaction mix.....	48
Table 2.7. TVR PCR cycling conditions.....	48
Table 2.8. 5p8:17p6:XpYpM:16p1:21q1 fusion PCR reaction mix.....	49
Table 2.9. Fusion PCR cycling conditions.....	49
Table 2.10. Fusion reamplification PCR reaction mix.....	50
Table 2.11. Sequencing PCR reaction mix.....	54
Table 2.12. Sequencing PCR cycling conditions.....	55
Table 3.1 Descriptive analysis at the 5p, 17p and XpYp telomeres for 57 CLL patients.	68
Table 3.2. <i>In-silico</i> PCR results using the 5p9 and 21q1 primers.	79
Table 3.3. Frequency of fusions detected with the 17p6:XpYpM:16p1:21q1 compared to the 5p8:17p6:XpYpM:16p1:21q1 fusion assay.....	83
Table 3.4. CLL patient samples selected to study telomere dysfunction and fusion.....	90
Table 4.1. Classification of type of event.....	116
Table 4.2. Validated telomere fusion events identified from a panel of 9 CLL patients.....	127
Table 4.3. Microhomology at the fusion junction.....	134
Table 4.4. Description of genomic loci that fuses with dysfunctional telomeres.....	142
Table 5.1. Clonal and subclonal validated mutations.....	173
Table 5.2. Transition/Transversion ratio.....	177
Table 5.3. VAF of 9 mutations on the <i>IGHV</i> ₄₋₆₁ locus for CLL patient DB17.....	182

LIST OF SUPPLEMENTARY FIGURES

S. Figure 1. STELA protocol for 6 samples.....	219
S. Figure 2. Telomere fusion protocol for 6 samples.....	220
S. Figure 3. Identifying SNVs present on the tumour sample.....	221
S. Figure 4. CNVs identified across the genome of DB17 CLL-B cells.....	222
S. Figure 5. Copy number deletion at chr6q21 and chr13q14.....	223

S. Figure 6. Sequencing the breakpoint A of the 6q:13q translocation	224
S. Figure 7. Sequencing the breakpoint B of the 6q:13q translocation	225

LIST OF SUPPLEMENTARY TABLES

S. Table 1. List of oligonucleotides used during this study.	226
S. Table 2. Characteristics of 5p STELA PCR primers.....	228
S. Table 3. Comparison of the mean telomere length (Kb) and standard deviation at 5p, 17p and XpYp for 57 CLL patient samples.....	228
S. Table 4. TL before and after correcting for TVR limit (Kb).	229
S. Table 5. Oligonucleotide sequences for the 5p chromosome end: distance to telomere, primer ID, sequence, length, expected T_m and GC content.....	230
S. Table 6. 5p primer combination and PCR product.....	230
S. Table 7. Telomere dynamics at the 5p chromosome end for HCT116 ^{DN-hTERT}	230
S. Table 8. Number and proportion of CLL patients for each category of fusions.....	231
S. Table 9. Description of potential 5p fusions detected in CLL patient samples.	232
S. Table 10. Summary of events resulting from intra-chromosomal fusion analysis.	233
S. Table 11. Summary of events resulting from inter-chromosomal fusion analysis.	234
S. Table 12. Summary of events resulting from intra/inter-chromosomal fusion analyses.....	235
S. Table 13. Type and proportion of TVRs identified.	236
S. Table 14. Biological process gene ontologies enriched in gene list, with validated fusion junctions, disrupted by inter-chromosomal fusion events.....	237
S. Table 15. List of genes with validated fusion junction submitted to DAVID and GSEA. ...	239
S. Table 16. GSEA MSigDB gene set collections (Mootha, 2003; Subramanian, 2005).	240
S. Table 17. Enrichment in GSEA MSig Collections from a list of disrupted genes with validated fusion junction	240
S. Table 18. Gene set overlaps found between a list of genes disrupted by inter-chromosomal telomere fusions and the MSigDB using GSEA.	241
S. Table 19. Summary of impact of clonal and subclonal SNVs detected in CLL-B cells	242
S. Table 20. High and moderate impact mutations in tumour CLL-B cells.....	243
S. Table 21. List of genes with subclonal mutations.	244
S. Table 22. Enrichment in gene ontology biological process from subclonal mutations. ...	246
S. Table 23. Clonal mutations are enriched in the MAPK cascade.....	249
S. Table 24. 7 mutations disrupting POT1-TPP1 binding.	249

CHAPTER 1:

INTRODUCTION

1.1 THE TELOMERE SYSTEM

The telomere was first discovered over 80 years ago by geneticist Hermann Muller when he was studying X-ray induced DNA breaks in *Drosophila melanogaster*. He observed that the chromosome ends were never involved in chromosomal rearrangements, leading him to hypothesise that the 'terminal gene' had a protective function (Muller, 1938). Not long after, Barbara McClintock converged on the same conclusion studying chromosomal aberrations resulting from ionizing radiation in *Zea mays*. She observed that when the terminal structures of DNA were lost, chromosome ends would fuse (McClintock, 1941).

Since then, it has been established that telomeres are nucleoprotein complexes at the end of linear chromosomes composed of the telomere-repeat DNA sequence (**section 1.1.1**) and the shelterin complex (**section 1.1.2**). In addition, telomere length (TL) is elongated by telomere maintenance mechanisms (**section 1.1.5**) in stem, germline and cancer cells but not in somatic cells (**Figure 1.1**).

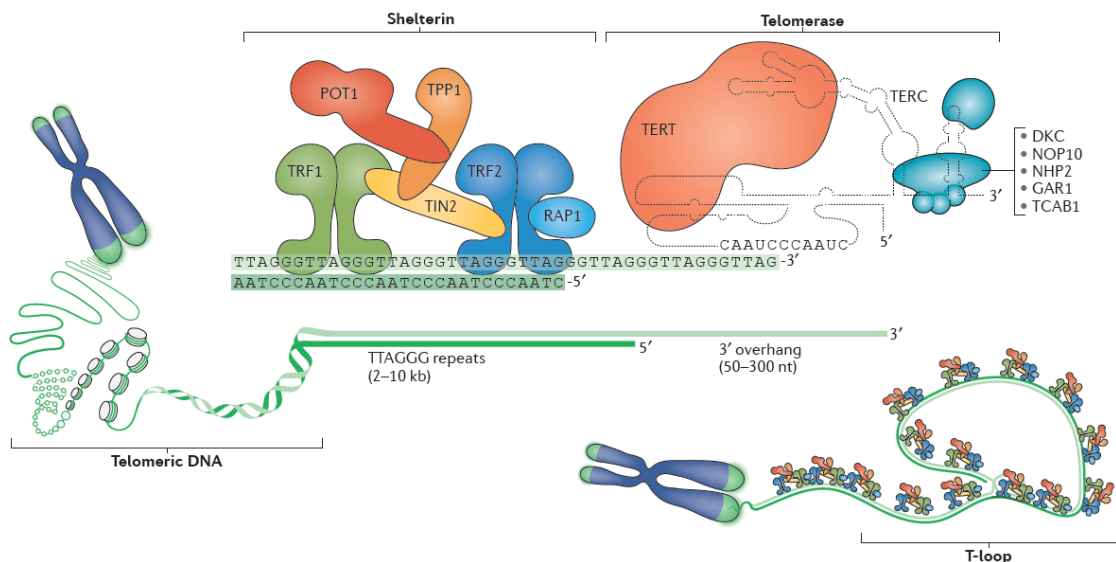


Figure 1.1. The human telomeric system.

The telomere system is composed of telomeric DNA sequence, the shelterin complex and the telomerase enzyme. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology (Maciejowsky and De Lange, 2017), copyright (2017)

The telomere system is conserved across the five kingdoms of life, particularly amongst eukaryotes regardless of minor differences in TL and the sequence repeat, highlighting the fundamental role they play in the cell.

Telomeres are implicated in cellular senescence, dictating the lifespan of the cell (**section 1.1.5.2**). Moreover, telomeres play an important role in maintaining genomic integrity by protecting chromosome ends from being mistaken for a double-strand break (DSB) (**sections 1.1.3 and 1.1.4**). Otherwise, dysfunctional telomeres can be subjected to enzymatic end-degradation, activation of DNA repair mechanisms and telomere fusion events that can initiate genomic instability and lead to malignant transformation (**section 1.2**).

1.1.1 The telomeric sequence

Telomeric DNA is characterised by double stranded short tandem repeats, with a G-rich single strand sequence that protrudes in an overhang at the 3' end and a shorter C-rich strand at the 5' end.

The human telomeric repeat is composed of TTAGGG and it was first identified in 1988 (Moyzis *et al.*, 1988). It can extend up to 20Kb and the 3' overhang contains 50-300 nucleotides depending on the individual, the age and the tissue samples examined (Baird *et al.*, 2006, Griffith *et al.*, 1999). Differences across cell types occur because of a differential replication rate and expression of telomerase that will determine telomere erosion (Griffith *et al.*, 1999, Baird *et al.*, 2006, Capper *et al.*, 2007, Lin *et al.*, 2010). In human somatic cells that do not express telomerase, telomeres shorten 60-120bp per cell division (Harley *et al.*, 1990, Baird *et al.*, 2003). Furthermore, there are differences in TL across species, for example: 0.25Kb of TTGGG in protist *Tetrahymena thermophile* (Blackburn and Gall, 1978, Jacob *et al.*, 2004), 2-9Kb of TTTAGGG in plants (Richards and Ausubel, 1988, Shakirov and Shippen, 2004) and 20-150Kb of TTAGGG in inbred laboratory mice (Kipling and Cooke, 1990, Hemann and Greider, 2000).

In humans, a degenerate telomere variant repeat (TVR) is also found at the start of the telomere (first 1-2Kb) (de Lange *et al.*, 1990, Baird, 1995, Coleman *et al.*, 1999, Baird *et al.*, 2003). The most common TVRs identified include TCAGGG, TGAGGG and TTGGGG, in addition to other less frequent variants identified in human cell lines (Letsolo *et al.*, 2010, Lee *et al.*, 2014a). TVRs have altered function from canonical repeats; telomeres lose their protecting capping function when they erode up to the TVR distribution and the variant

CTAGGG was shown to decrease the binding affinity of the shelterin members TRF1 and TRF2 (Capper *et al.*, 2007, Mendez-Bermudez *et al.*, 2009).

1.1.2 The Shelterin complex

To maintain its protective function, the telomeric sequence is assisted by the shelterin complex. Shelterin, also termed telosome, is a 6-subunit telomere-associated protein complex that consist of telomeric repeat-binding factor 1 (TRF1), telomeric repeat-binding factor 2 (TRF2), protection of telomeres 1 (POT1), TRF1- and TRF2-interacting nuclear factor 2 (TIN2), Repressor/Activator Protein 1 (RAP1) and TINT1-PTOP-PIP1 (TPP1) (de Lange, 2005a).

POT1 binds to the telomeric 3' single-stranded G-overhang and to TPP1, which is associated to TIN2. TIN2 acts as a bridge and connects TRF1 and TRF2, which bind directly to double-stranded telomeric DNA recruiting the shelterin complex to telomeres. RAP1 is associated with TRF2 (deLange, 2005). Hundreds of shelterin complexes are present at every telomere in each cell (de Lange, 2005a) (**Figure 1.1**).

1.1.3 Telomere protection: hiding the end and inhibiting DSB response

Aided by the shelterin complex, the telomeric G-strand 3' overhang loops back to form a protective T-loop structure by invading the dsDNA and pairing with the C-rich strand which displaces the G-strand and also creates a small D-loop (**Figure 1.2B**) (Griffith *et al.*, 1999, Greider, 1999). This process helps to sequester the 3' overhang to stabilise the telomere and prevent it from triggering the DNA damage response and recruiting the DNA repair machinery (de Lange, 2005a).

Additionally, the single-stranded telomeric G-rich overhang can also form G-quadruplex structures (**Figure 1.2C**) - secondary structures created by the interaction of four guanine bases stabilised by hydrogen bonding and a central cation in square planar arrangements that stack on top of each other (Williamson *et al.*, 1989, Biffi *et al.*, 2013). G-quadruplexes contribute to telomere protection by sequestering the 3' overhang (Smith *et al.*, 2011).

It has also been suggested that telomeres may also become inaccessible to DNA damage response (DDR) components by a shelterin-mediated 10-fold compaction of telomeric chromatin, for which each shelterin subunit is necessary (Bandaria *et al.*, 2016).

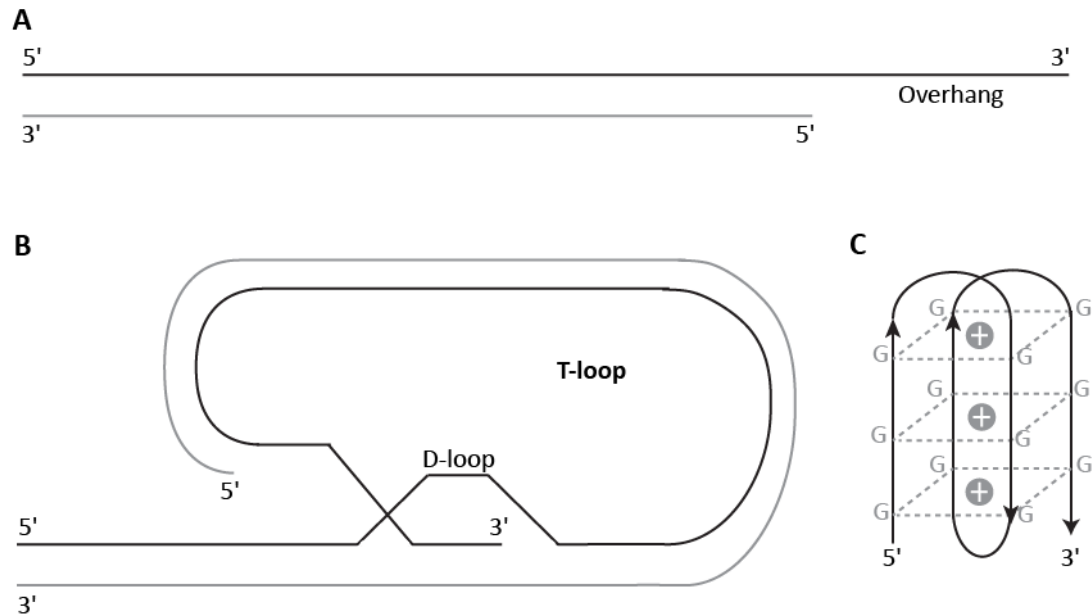


Figure 1.2. The structure of telomeric DNA.

Telomeric DNA in a (A) linear conformation, (B) T-loop structure and (C) G-quadruplex structure.

The shelterin members TRF1, TRF2, RAP1 and POT1/TPP1 inhibit the DDR. TRF2 and POT1 modulate the ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3 related (ATR) pathways respectively, protecting telomeres from being recognised as DSB (Karseder *et al.*, 2004, Liu *et al.*, 2004, Martinez *et al.*, 2009, Sfeir *et al.*, 2010, Thanasoula *et al.*, 2012). By repressing the DDR they block the homologous recombination (HR), classical and alternative non-homologous end joining (C/A-NHEJ) DNA repair mechanisms at telomeres that could result in telomere fusions (reviewed in **section 1.1.4**). Therefore, telomere deprotection by deletion of either TRF2 or POT1 has been associated with C-NHEJ-mediated telomere fusions that do not depend on the TL (Yang *et al.*, 2005, Celli and de Lange, 2005). However, shelterin does not suppress the DDR at induced-DSBs within the internal telomeric repeat array, that are repaired by HR and A-NHEJ, most likely to maintain telomere integrity (Doksani and de Lange, 2016).

1.1.4 DNA damage response and DNA repair mechanisms at telomeres

Maintaining genomic stability is crucial for cell survival. DSBs are extremely cytotoxic DNA lesions that if left unrepaired can cause cell death. However, faulty repair mechanisms can create chromosomal rearrangements that lead to tumorigenesis (Jackson and Bartek, 2009).

In the presence of a DSB, the MRE11–RAD50–NBS1 (MRN) complex locates to it and recruits the ATM kinase which becomes active and phosphorylates the histone variant H2AX. Phosphorylated H2AX recruits other proteins including p53-binding protein 1 (53BP1) and Rap1-interacting factor 1 (RIF1) to DNA damage foci that amplify the DNA damage signal and regulate the repair, mediated by C-NHEJ (Riballo *et al.*, 2004, Escribano-Diaz *et al.*, 2013).

DSB can also be resected 5'-3' by the MRN complex, the CtBP-interacting protein (CtIP) and Breast Cancer Type 1 Susceptibility protein (BRCA1), generating a tail of ssDNA that is recognised by the single-strand-binding replication protein A (RPA) which recruits ATR (Xie *et al.*, 2009, Choi *et al.*, 2010, Badie *et al.*, 2015, Anand *et al.*, 2016). Activated ATR, similarly to ATM, triggers the DNA damage signalling and initiates repair (Brown and Baltimore, 2003). Resection to reveal microhomology is necessary for homology directed repair (HDR) mechanisms including alternative NHEJ (A-NHEJ), single-strand annealing (SSA) and homologous recombination (HR).

C-NHEJ and HR are the major DSB repair mechanisms but recently, A-NHEJ and SSA have also been implicated. The distinct mechanisms and the cell cycle effects are explained in the following sections and summarised in **Figure 1.3**.

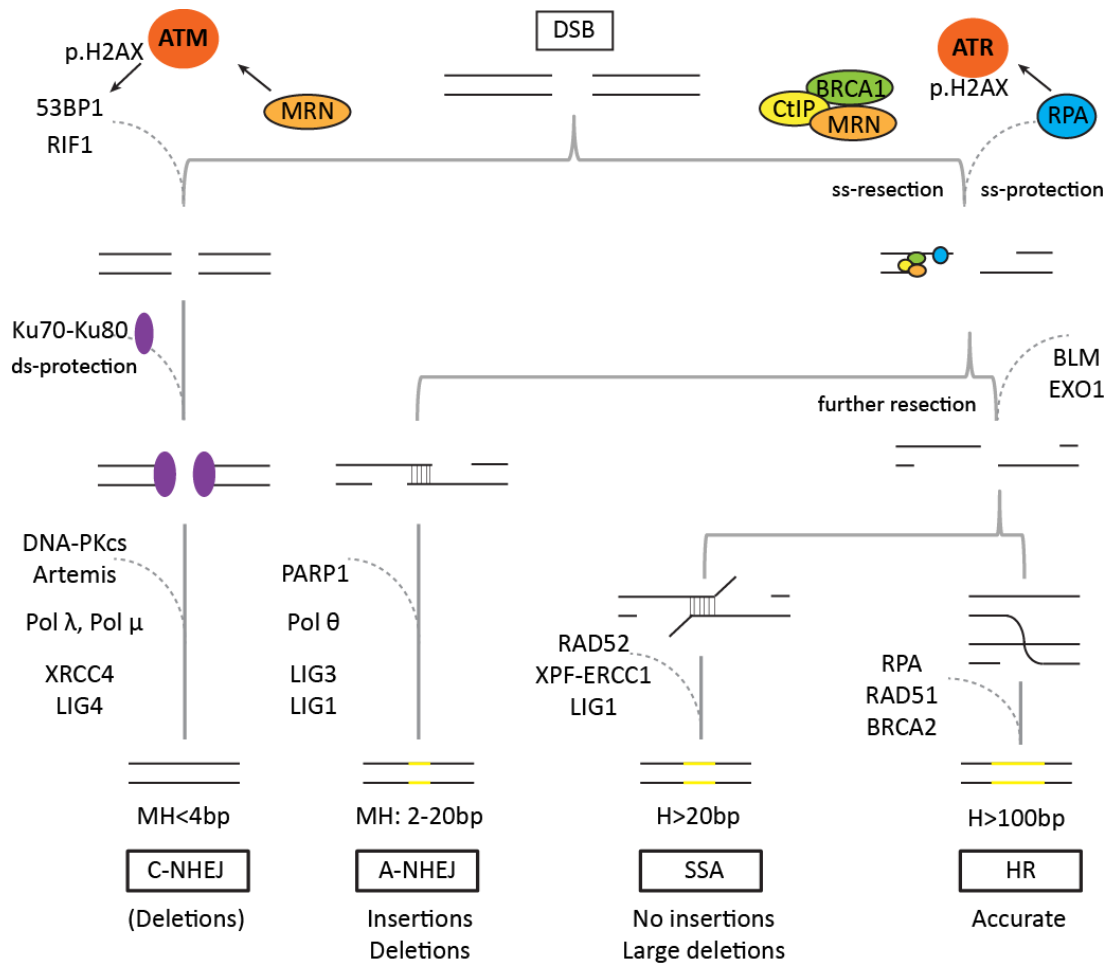


Figure 1.3. Synthesis of DSB and SSB repair mechanisms.

A double-strand break (DSB) can be repaired by the classical non-homologous end joining (C-NHEJ), alternative non-homologous end joining (A-NHEJ), single-strand annealing (SSA) or homologous recombination (HR) pathways. The pathway choice will depend on the cell cycle stage, the availability of the different molecules involved, the end resection and the usage of (micro) homology. Genetic outcomes that can result from each pathway and usage of microhomology (MH) or homology (H) are stated. Figure adapted from Chang *et al.*, 2017.

1.1.4.1 Classical non-homologous end joining (C-NHEJ)

In G1 cells, DSBs can be repaired by resection-dependent or resection-independent C-NHEJ (Biehs *et al.*, 2017). DSBs are sensed by the Ku70-Ku80 heterodimer (Ku) that protect the ends from extensive resection and serves as a platform to recruit other proteins (Mimitou and Symington, 2010). Ku forms a complex with DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and binds to another Ku-DNA-PKcs complex at the other end of the break acting like a bridge and bringing the DNA ends together using little or no sequence

homology (Weterings and Chen, 2008). DSB ligation is catalysed by a complex composed of DNA ligase 4 (LIG4) and X-ray repair cross-complementing protein 4 (XRCC4) (Lieber, 2010). Depending on the DNA end configuration (blunt, incompatible, resection-dependent compatible ends), various NHEJ sub-pathways with different proteins involved have been suggested. For example, Artemis can reveal small microhomology (≤ 4 nt) regions after resection and cleave the ss-overhangs; and when ends are incompatible, DNA polymerase μ (Pol μ) and Pol λ are involved in the DNA synthesis in a template-dependent/independent manner (Chang *et al.*, 2016, Chang *et al.*, 2017). Ionizing radiation (IR)-induced DSBs at heterochromatin or regions with additional damage, potentially when repair is delayed, Polo-like kinase 3 (Plk3) phosphorylates CtIP that binds BRCA1 and promotes resection by exonucleases EXO1, EXD2 and MRE11. The process is completed by endonuclease Artemis and repaired by resection-dependent C-NHEJ. This pathway could also represent microhomology-mediated end joining (MMEJ) and can result in deletions and translocations (Barton *et al.*, 2014, Biehs *et al.*, 2017).

1.1.4.2 Alternative non-homologous end joining (A-NHEJ)

A-NHEJ, often taken to be synonymous of MMEJ, has historically been considered an error-prone repair mechanism together with C-NHEJ. During G2, A-NHEJ is activated when DNA damage at the resected end is detected by Poly (ADP-ribose) polymerase 1 (PARP1) (Robert *et al.*, 2009). Typically 2-20bp (most common 4-6bp) of microhomology is annealed and low-fidelity DNA Pol θ stabilises the ssDNA overhangs, which is then ligated by DNA LIG3 or LIG1 (Mateos-Gomez *et al.*, 2015, Ceccaldi *et al.*, 2015, Kent *et al.*, 2015, Chang *et al.*, 2017). If additional microhomology is required, Pol θ can also add nucleotides to the ends, including templated insertions (Wyatt *et al.*, 2016).

1.1.4.3 Single-strand annealing (SSA)

Greater homology is required for the annealing of complementary ssDNA tails during repair by SSA and HR; thus, further resection is performed by BLM and EXO1 during S/G2 phases (Tomimatsu *et al.*, 2014, Symington, 2016). Then RPA, which is crucial for ATR activation, binds to ssDNA tails to facilitate repair (Zou and Elledge, 2003). For SSA, RAD52 mediates alignment of large homologous sequences (> 20 bp) and the XPF-ERCC1 complex excises the remaining DNA overhangs before ligation, leading to large deletions (Bhargava *et al.*, 2016).

1.1.4.4 Homologous recombination (HR)

For the HR pathway, a sister or homologous chromatid is essential for repair. Over 100bp of homology is required and there is no loss of nucleotides (Moynahan and Jasin, 2010). Strand invasion and homology search is mediated by RAD51 that binds to the ssDNA tail (Sung, 1994, Sung *et al.*, 2003). HR generally performs a precise repair of sister chromatids when DSBs are generated during S and G2 phases, when sister chromatids are nearby. However, it can result in telomere sister chromatid exchanges, telomere-loop excision and alternative lengthening of telomeres (ALT) (**section 1.1.5.4**) (Cho *et al.*, 2014).

1.1.4.5 Choice of pathway and implications in cancer

The DNA repair pathway choice across the genome depends on the extent of end-resection from the DSB and is heavily influence by the cell cycle phase. C-NHEJ is active throughout the cell cycle, particularly during G1 phase, when 53BP1 and RIF1 block recruitment of BRCA1 and inhibit end-resection from the DSB. However, a resection-dependent C-NHEJ also operates during G1 (Biehs *et al.*, 2017). During S/G2 phase, cyclin-dependent kinases (CDKs) phosphorylate the MRN complex and CtIP stimulating extensive end-resection and, in addition with BRCA1, promote DSB repair by A-NHEJ, SSA and HR. Then, the usage of (micro) homology will mainly determine the pathway choice (Ceccaldi *et al.*, 2016).

It has been suggested that C-NHEJ is more commonly used than A-NHEJ due to Ku having a higher binding affinity to the DSB than PARP1 and MRN (Wang *et al.*, 2006, Mimitou and Symington, 2010, Cheng *et al.*, 2011). In addition, together with 53BP1 and the DNA Pol zeta (ζ) processivity subunit REV7, Ku also represses DNA end resection at DSBs obstructing the use of the alternative mechanism (Lottersberger *et al.*, 2013, Boersma *et al.*, 2015). DSB repair by resection-dependent C-NHEJ can lead to deletions and translocations (Barton *et al.*, 2014, Biehs *et al.*, 2017). Furthermore, C-NHEJ has been proposed to mediate the ligation of the DNA fragments resulting from chromosome shattering (chromothripsis) which leads to a highly rearranged chromosome (chromoanagenesis) that can be found in human tumours (Holland and Cleveland, 2012).

When sister chromatids are not available for HR, repair is mediated by A-NHEJ leading to deletions, translocations and chromosomal rearrangements (Zhang and Jasin, 2011, Ceccaldi *et al.*, 2016). It has also been demonstrated that when C-NHEJ is not functional or both ATR and ATM are activated, A-NHEJ is used (Rai *et al.*, 2010, Sfeir and de Lange, 2012). Moreover, when HR is compromised, Pol θ may promote the use of the A-NHEJ mechanism

by binding and inhibiting RAD51 (Simsek *et al.*, 2011, Mateos-Gomez *et al.*, 2015). Interestingly, Pol θ was upregulated in HR-deficient breast and ovarian cancers (Mateos-Gomez *et al.*, 2015, Ceccaldi *et al.*, 2015), and was associated with cellular transformation and poor prognosis (Lemee *et al.*, 2010, Higgins *et al.*, 2010). Generation of human chromosomal translocations arising from genomic DSBs have been mainly attributed to C-NHEJ (Ghezraoui *et al.*, 2014, Lieber, 2016). However, more studies are revealing the implication of A-NHEJ in cancer (Simsek *et al.*, 2011).

1.1.4.6 DNA repair activity at telomeres

Since unprotected telomeres are recognised and processed as DSBs, this section aims to highlight the specific repair mechanisms reported at telomeres and whether their repair differs from non-telomeric DSBs.

A deficient processing of subtelomeric compared to interstitial DSBs has been observed in the EJ-30 human bladder carcinoma cell line after induction of DSBs. Repair at the subtelomere was characterised by large resection, increased frequency of deletions and chromosomal rearrangements. This was also observed at interstitial sites containing telomeric DNA (Zschenker *et al.*, 2009). The different repair choice at telomeric DSBs was not a result from deficient C-NHEJ or HR but instead, it was related to excessive processing by nucleases including MRE11 at G1-phase. (Miller *et al.*, 2011, Muraki *et al.*, 2015, Alcaraz Silva *et al.*, 2017). Despite the mechanism responsible was not clear, Alcaraz Silva and colleagues proposed a role for A-NHEJ.

Fusion of short dysfunctional telomeres can be mediated by both LIG3/LIG1 A-NHEJ and LIG4 C-NHEJ pathways. A-NHEJ leads primarily to intra-chromosomal sister chromatid fusions but also to inter-chromosomal fusions in low proportion, and is characterised by greater resection and usage of microhomology at the fusion junction. At intra- and inter-chromosomal junctions a mean of 2.9nt and 2.38nt of microhomology respectively was observed in HCT116 LIG4^{-/-} colorectal cancer cells. In contrast, C-NHEJ resulted in inter-chromosomal fusions with non-telomeric loci and with telomeric-repeat regions with reduced microhomology usage (Jones *et al.*, 2014, Liddiard *et al.*, 2016). Furthermore, a role of LIG1 in A-NHEJ has been inferred in the absence of LIG3 or LIG4 (Arakawa *et al.*, 2012, Lu *et al.*, 2016, Liddiard *et al.*, 2016). Sequence analysis of telomere fusions in human cancer cells revealed extensive resection and microhomology usage at the fusion junction consistent with A-NHEJ (Letsolo *et al.*, 2010, Lin *et al.*, 2010).

The frequency of fusions increases when the tumour-suppressor gene *TP53* is disrupted (Artandi *et al.*, 2000, Liddiard *et al.*, 2016), explained by the loss of genomic integrity when telomeres become dysfunctional and this cell cycle checkpoint is lost. Furthermore, it could also be related to the loss of p53-associated subtelomeric protection, a recently identified function (Tutton *et al.*, 2016). In addition to chromosome end-to-end fusions, dysfunctional telomeres can fuse with non-telomeric loci, primarily within coding regions. Gene disruption most likely has harmful consequences for cell viability but may also provide a mechanism for malignant transformation (Liddiard *et al.*, 2016).

In HCT116 cells expressing dominant negative telomerase (DN-*hTERT*) to induce a telomere-driven crisis, LIG3 was vital to clonal evolution and escape. These results suggested that sister-chromatid fusion events that lead to amplification or deletion of chromosome arms may provide a selective advantage to initiate tumorigenesis compared with LIG4-mediated long-range inter-chromosomal fusions (Jones *et al.*, 2014). For some clones, escape from crisis was characterised by amplification of the 5p chromosome arm that harbours *hTERT*, concurrent with reactivation of telomerase (Jones *et al.*, 2014).

Altogether, telomeric DSBs can be repaired by the NHEJ pathways resulting in intra- and inter-chromosomal fusions with different microhomology usage, deletions and insertions. This is consistent with the mutational signature observed at the telomere fusion junctions in human cancers cells. Thus, C-NHEJ and A-NHEJ-mediated repair of dysfunctional telomeres may be responsible for many of the gross chromosomal rearrangements detected in human malignancies.

1.1.5 Telomere maintenance mechanisms

1.1.5.1 Telomerase

Telomerase is the ribonucleoprotein complex that elongates the telomere via reverse transcription by adding TTAGGG repeats to the 3' terminus as it recognises the hydroxyl group (OH) (Greider and Blackburn, 1985, Morin, 1989). Telomerase expression is stable in germline and stem cells, allowing them to maintain the TL, but low or undetectable in somatic cells, which results in telomere shortening (Harley *et al.*, 1990, Baird *et al.*, 2003). Moreover, reactivation of telomerase is required for immortalisation and malignant progression; by lengthening the telomeres it expands cell replicative the lifespan (Counter *et al.*, 1992, Kim *et al.*, 1994).

The telomerase complex consists of 3 subunits: the catalytic subunit telomerase reverse transcriptase (TERT), the telomerase RNA template component (TERC) and accessory proteins. The telomerase accessory proteins include dyskerin (DKC), nuclear protein 10 (NOP10), non-histone protein 2 (NHP2), GAR1 and telomerase Cajal body protein 1 (TCAB1) (Egan and Collins, 2010). The human *TERT* gene is located at the end of chromosome 5p (Chr5p15.33; Chr5: 1,253,262-1,295,184; CRGCh37), 1.2Mb from the telomere and is approximately 40Kb long. TERT contains four domains: the N-terminal extension (TEN) domain with affinity to telomeric ssDNA and TPP1; the telomerase RNA binding (TRB) domain that interacts with TERC; the reverse transcriptase (RT) domain responsible for the catalytic activity and the C-terminal extension (CTE) domain (Sandin and Rhodes, 2014). The telomerase RNA component, *TERC*, is located at Chr3q26.2 (Chr3: 169,482,308-169,482,848) and is approximately 500bp long. It is composed of the catalytically essential pseudoknot-template core (PK) domain and the CR4/CR5 stem-loop (STE) domain that interacts with TERT, and the H/ACA scaRNA domain that binds telomerase accessory proteins (Blackburn and Collins, 2011, Zhang *et al.*, 2011a).

TCAB1 promotes assembly of the telomerase complex in the Cajal body (Stern *et al.*, 2012) and, together with TPP1, recruits it to the telomere (Tejera *et al.*, 2010, Nandakumar and Cech, 2013). Telomerase aligns its RNA templated sequence (AATCCC) to the telomeric 3' overhang and elongates the G-rich DNA by adding *de novo* TTAGGG repeats. Then the RNA template dissociates from the DNA and translocates along the telomere to repeat the process (**Figure 1.4**). Finally, a primer is synthesised at the complementary strand and DNA is replicated completing elongation of telomeres (Greider and Blackburn, 1989, Morin, 1989, Blackburn and Collins, 2011).

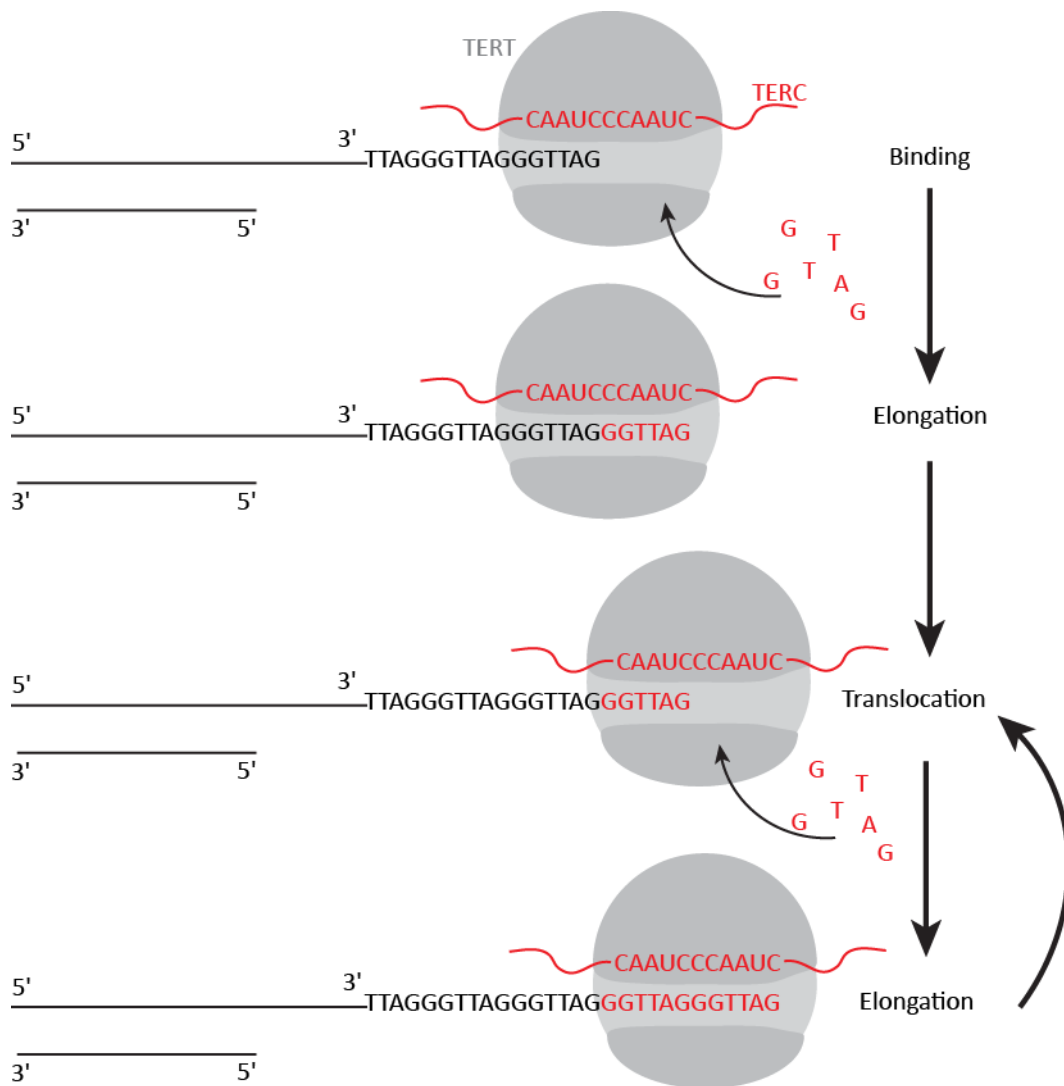


Figure 1.4. Telomere length elongation by telomerase.

The telomerase complex binds to the 3' overhang and using the RNA template TERC, the catalytic subunit TERT synthesises the DNA. Then, the complex translocates towards the end and the process is repeated.

1.1.5.2 Telomere shortening and senescence: a tumour suppressor mechanism

30 years after the discovery of telomeres, A. Olovnikov and J. D. Watson noted that the somatic mammalian cell was unable to reproduce a complete copy of linear chromosomes. This resulted in shorter telomeres every cell division, referred to as marginotomy or the end-replication problem (Olovnikov, 1971, Watson, 1972). Thus, telomeres protect the non-telomeric chromosomal coding sequence from being eroded when the genome is replicated ahead of cell division (**Figure 1.5**).

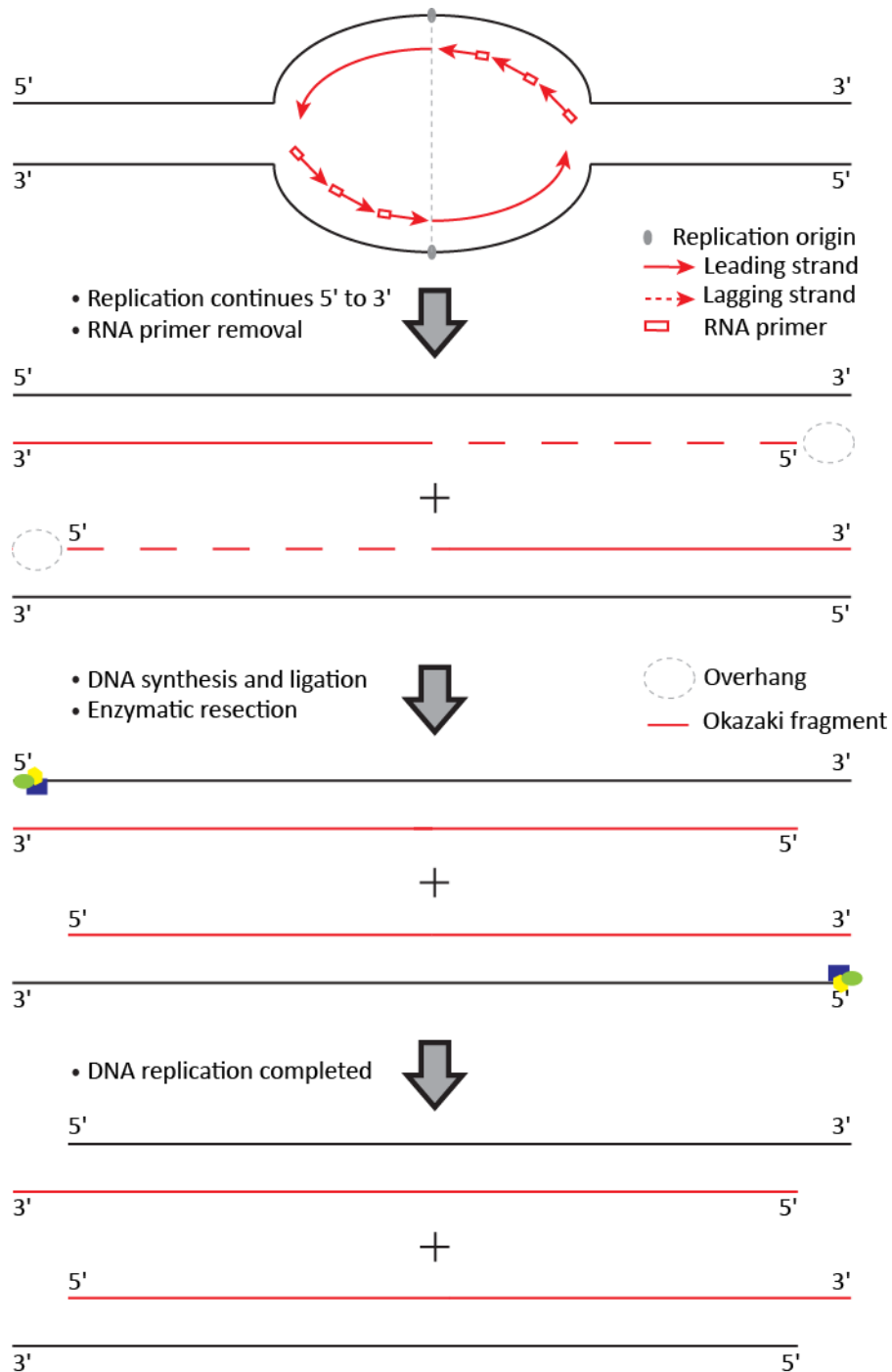


Figure 1.5. The end-replication problem.

Helicases unwind the dsDNA at the replication origin. DNA is synthesised in a 5' to 3' direction resulting in a leading and a lagging strand. The later one is synthesised discontinuously since it requires RNA primers to start replication. Then RNA primers are removed resulting in small fragments of DNA called Okazaki fragments and a 3' overhang. DNA is synthesised and ligated, and enzymatic resection leads to the formation of the 3' overhang in the copied strand.

DNA polymerase replicates genomic DNA in a 5' to 3' direction, which results in a different copying system for the leading and the lagging strand characterised by a continuous and discontinuous synthesis, respectively. To copy the lagging strand, RNA primers are required as the replication fork moves outwards from the origin of replication, resulting in short synthesised DNA fragments called Okazaki fragments. The internal RNA primers are then replaced by their DNA counterparts and ligated together to create a continuous strand; excluding the terminal primer for which there is no available 3' OH group to add nucleotides to (Ohki *et al.*, 2001). The loss of nucleotides at the chromosome end of the lagging strand leads to the G-rich ss-overhang at the telomere. However, since the leading strand results in blunt DNA ends, the 5' DNA terminus needs to be further resected to maintain the 3' overhang (Huffman *et al.*, 2000). This occurs by exonucleolytic enzymes Apollo1 and EXO1 that digest the 5' terminus, followed by a CST-associated Pol alpha (α)-mediated synthesis step (Huffman *et al.*, 2000, Chai *et al.*, 2006, Dai *et al.*, 2010, Wu *et al.*, 2012) (**Figure 1.5**).

Olovnikov likened the continuous shortening of the 'telogene' after every somatic cell division to an inner clock that would allow the cell to divide a specific number of times before reaching a critical length that would trigger senescence (Olovnikov, 1971, Olovnikov, 1973). Olovnikov's idea explained the intrinsic cell division counting mechanism previously observed by Hayflick and Moorhead (Hayflick and Moorhead, 1961, Hayflick, 1965).

Altogether, in human somatic cells that do not express telomerase, telomeres shorten by 50-100bp per population doubling (Harley *et al.*, 1990, Hodes, 1999, Baird *et al.*, 2003) and telomere shortening is considered a principal ageing mechanism that determines the lifespan of cells, known as the Hayflick limit. However, this loss can be neutralized with the extension of TL as evidenced with the increase of cell longevity following addition of telomerase (Bodnar *et al.*, 1998). This is also observed in germline, stem and cancer cells that express telomerase.

When telomeres become critically short or uncapped by shelterin depletion, they become unprotected and a DNA-damage response is triggered via ATM/ATR signalling. This culminates in the activation of replicative senescence that ceases cell division arresting cells in G₁/S-transition (Fabrizio d'Adda di Fagagna *et al.*, 2003, Artandi and Attardi, 2005). Moreover, a few critically short telomeres are sufficient to trigger the DDR (Hermann, 2001). This is a tumour suppressor mechanism that protects organisms from the propagation of abnormal cells with accumulated DNA damage over time that could otherwise result in tumour development (see **section 1.2.1**).

1.1.5.3 Telomere length regulation by modulating telomerase

The shelterin complex is implicated in regulating TL by modulating the access of telomerase (Palm and de Lange, 2008, Greider, 2016). TPP1 recruits TERT to telomeres for telomere elongation, consequently TPP1-depletion leads to telomere shortening and decreased TERT binding to telomeres (Tejera *et al.*, 2010). TRF1 is a negative regulator of telomerase-dependent telomere replication; overexpression of TRF1 results in telomere shortening while TRF1-depletion promotes telomere lengthening (van Steensel and de Lange, 1997, Ancelin *et al.*, 2002, Ho *et al.*, 2016). In addition, TL regulation may be assisted by POT1 through its binding to telomeric ssDNA (Colgin *et al.*, 2003, Loayza *et al.*, 2004).

TPP1 has also been implicated in TL regulation. This could be via its interaction with POT1, or through a telomerase-interacting TPP1 glutamate (E) and leucine (L)-rich (TEL) patch required for the recruitment of telomerase to telomeres and for stimulating telomerase-mediated addition of multiple telomeric repeats (Nandakumar *et al.*, 2012, Zhong *et al.*, 2012).

Indirectly, TIN2 contributes to TL regulation by modifying tankyrase 1, a telomeric poly(ADP-ribose) polymerase (PARP), that in turns inhibits the binding of TRF1 to the telomere. Inhibition of TIN2 leads to decreased TRF1 binding and results in telomere elongation (Smith and de Lange, 2000, Ye and de Lange, 2004).

Recently, another telomere-associated protein has been proposed to modulate TL: telomeric zinc finger-associated protein (TZAP). It is proposed that TZAP competes with TRF1 and TRF2 to bind telomeres and therefore preferentially localises to long telomeres, that have a low density of shelterin, and trims telomeric repeats, preventing aberrantly long telomeres (Li *et al.*, 2017).

1.1.5.4 Alternative lengthening of telomeres (ALT)

ALT is a telomerase-independent telomere maintenance mechanism that has been observed in a proportion of immortalised cells in culture and 10-15% of human cancers (Reddel, 2000, Reddel *et al.*, 2001, Henson *et al.*, 2002) with greater enrichment in mesenchymal tumours (Henson *et al.*, 2005, Dilley and Greenberg, 2015).

ALT is based on homologous recombination (HR)-mediated DNA replication. The templates proposed for telomere lengthening in human cells include using a sister-chromatid telomere, copying itself through T-loop formation or looping out, or inter-chromosomal telomere copying (**Figure 1.6**) (Dunham *et al.*, 2000, Cho *et al.*, 2014, Pickett and Reddel,

2015). The steps required possibly consist of strand invasion creating a Holliday junction (HJ), copying the template, resolution of the HJ and synthesis of the complementary strand (Pickett and Reddel, 2015).

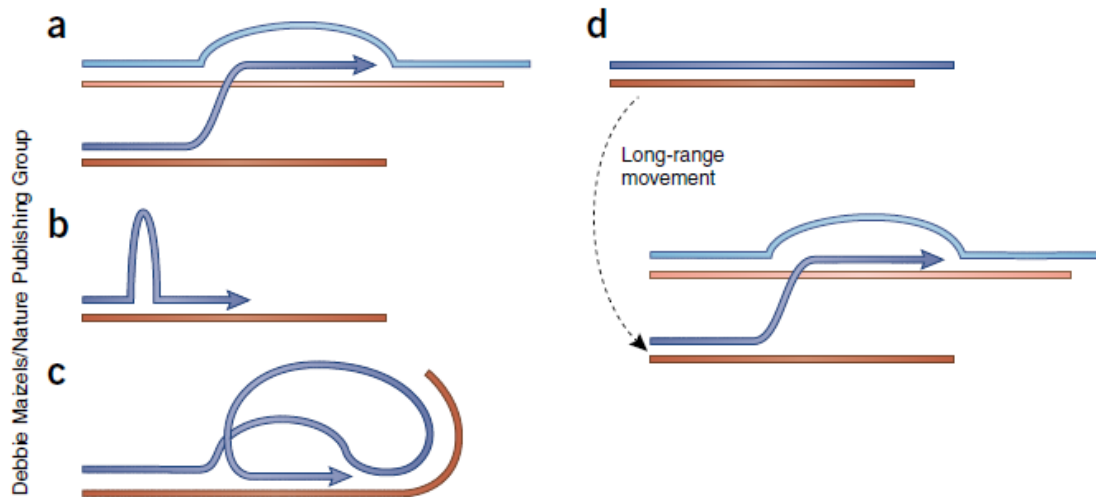


Figure 1.6. Templates for HR in ALT.

Homologous recombination with (a) a sister chromatid, the same telomere by (b) looping out or (c) t-loop formation, or (d) a distant telomere. Reprinted by permission from Macmillan Publishers Ltd: Nature Structural and Molecular Biology (Pickett and Reddel, 2015), copyright (2015).

Cells utilising ALT are characterised by heterogeneous and very long telomeres, the presence of linear or circular (T-circles) extrachromosomal telomeric repeats (ECTR) and ALT-associated promyelocytic leukaemia (PML) nuclear bodies (APBs) (Murnane *et al.*, 1994, Tokutake *et al.*, 1998, Cesare and Griffith, 2004). Telomeric DNA, binding proteins and machinery for DNA replication, synthesis and recombination have been observed within APBs (Yeager *et al.*, 1999). In addition, ALT cells exhibit high levels of telomere sister chromatid exchanges (T-SCEs) (Londono-Vallejo *et al.*, 2004).

HR-copying in ALT cells can lead to the amplification of TVRs throughout the telomere, disrupting the shelterin binding sites (Lee *et al.*, 2014a). This results in increased DDR which may stimulate HR repair pathway increasing ALT activity (Flynn *et al.*, 2015). In addition, the spread of TVRs can create *de novo* binding sites for proteins such as the orphan nuclear receptor NR2C/F with affinity to TCAGGG. This can lead to a bridge-interaction between the telomeric-genomic NR2C/F binding sites that can drive the insertion of telomeric sequence resulting in fragile sites across the genome prone to breakage and translocations. Altogether, it is a mechanism for telomere-driven genomic instability in cells that engage ALT (Conomos *et al.*, 2012, Marzec *et al.*, 2015).

1.1.6 Subtelomeric DNA, gene families, TERRA and telomere-position effect

Subtelomeric regions constitute the transition between the telomere repeat tract DNA and the chromosome-specific genomic sequence and range from 10-300Kb in size. Most human subtelomeric regions have been mapped resulting in a subtelomeric reference sequence (Riethman *et al.*, 2004, Ambrosini *et al.*, 2007, Stong *et al.*, 2014). Subtelomeres have been described as mosaics of multichromosomal blocks since they are made up of various repeated elements, although the composition differs between them (Mefford and Trask, 2002). Subtelomeres contain degenerate telomeric TTAGGG repeats, DNA unique to specific chromosome ends and segmental duplications known as subtelomeric repeat elements (SREs). SREs are mainly found at the distal region and share similarities between the different chromosome ends (Mefford and Trask, 2002, Riethman *et al.*, 2004, Riethman *et al.*, 2005). It is proposed that SREs are generated from translocations involving distinct chromosomes and sister chromatid exchanges (SCEs) indicating that human subtelomeres are hotspots for recombination events (Linardopoulou *et al.*, 2005, Rudd *et al.*, 2007).

The recombination that occurs at subtelomeres makes them dynamic and variable regions. Since the subtelomere harbours some gene families, recombination contributes to phenotypic diversity, ultimately helping organisms to adapt to the changing environment (Trask *et al.*, 1998, Freitas-Junior *et al.*, 2000). Examples of gene families in humans include the olfactory receptors (OR) (Trask *et al.*, 1998, Mefford *et al.*, 2001), the Wiscott-Aldrich Syndrome Protein family (*WASH*) (Linardopoulou *et al.*, 2007), zinc finger-containing genes like MZF-1 (Hoffman, 1996) and the immunoglobulin heavy-chain genes at 14q (Cook *et al.*, 1994). In *Plasmodium falciparum*, the protozoan parasite that causes malaria in humans, the location of virulence gene families at the subtelomere contributes to the diversity of antigens and evasion of the host's immune system (Freitas-Junior *et al.*, 2000).

Subtelomeric DNA is transcribed into a long non-coding RNA known as Telomeric Repeat-containing RNA (TERRA) that contains G-rich telomeric repeats (Azzalin *et al.*, 2007). TERRA has an essential role in telomere maintenance and cell survival. It can regulate telomerase activity by binding the RNA template and assist capping the chromosome ends by allowing POT1 to bind ssDNA and displacing RPA. Downregulation of TERRA activates the DDR at telomeres, and TERRA may also participate in the formation of heterochromatin at telomeres (Azzalin and Lingner, 2015, Cusanelli and Chartrand, 2015, Montero *et al.*, 2016).

Other subtelomeric elements are binding sites for CTCF and cohesion that are involved in the regulation of TERRA transcription (Deng *et al.*, 2012). These factors are also implicated in telomere protection since experimental depletion of CTCF or cohesion resulted in telomere-induced DNA damage foci (TIF) formation and destabilised the binding of TRF1 and TRF2 at the proximal telomeric repeats (Deng *et al.*, 2012, Stong *et al.*, 2014).

Human subtelomeric regions sharing high homology between different chromosome ends have been termed telomeric families (Brown *et al.*, 1990). These include the 16p family (16p, 1p, 9p, 12p, 15q, XqYq and the 2q14 interstitial locus) and 21q family (21q, 1q, 2q, 5q, 6q, 6p, 8p, 10q, 13q, 17q, 19p, 19q, 22q and the 2q13 interstitial locus) (Letsolo *et al.*, 2010). In contrast, the presence of unique subtelomeric regions at distinct chromosome ends has been very valuable for measuring the length of individual telomeres using Single Telomere Length Analysis (STELA) (Baird *et al.*, 2003). Altogether, sequencing subtelomeric DNA also allowed the development of a telomere fusion PCR assay that targets single chromosome ends and subtelomeric families (Capper *et al.*, 2007). To investigate telomere fusion events with the PCR fusion assay, oligonucleotides need to be designed at the subtelomeric region since critically short and unprotected telomeres can be resected into the subtelomeric sequence (Capper *et al.*, 2007). Both techniques are discussed further in **sections 1.2.3.5 and 1.2.3.6.**

At subtelomeres, a phenomenon known as telomere-position effect (TPE) that was originally identified in yeast, has been described in human cells (Baur *et al.*, 2001). It results in the epigenetic silencing of genes proximal to telomeres leading to decreased gene expression by modifying the heterochromatin status depending on the TL. It is proposed that subtelomeres may buffer the TPE transcriptional silencing from genes located proximal to the chromosome end (Mefford and Trask, 2002). Recently, research has shown that TL could influence the transcription of genes adjacent to the subtelomere through a looping mechanism that locates long telomeres into proximity of genes up to 10Mb away; it is known as TPE over long distance (TPE-OLD) (Robin *et al.*, 2014). It has been suggested that the human telomerase gene *hTERT*, that is proximal to the 5p telomere, could be regulated by a TPE-OLD mechanism. Long telomeres appeared to epigenetically repress *hTERT* and associated with a telomeric loop. In addition, this loop was disengaged and the chromatin status changed when telomeres became short (Kim *et al.*, 2016). It is possible that TRF2 and TERRA are also required for the maintenance of the telomeric loop (Kim *et al.*, 2016). Importantly, this mechanism could have implications in reactivation of telomerase in cancer cells with critically short telomeres.

1.2 TELOMERES AND TELOMERASE IN CANCER

1.2.1 Telomere crisis: a source of genomic instability and malignant transformation

For malignant transformation to occur cells need to overcome the mortality stages 1 (M1) and 2 (M2) (Wright *et al.*, 1989, Shay *et al.*, 1991). Cells can escape senescence (M1) by inactivating cell cycle checkpoints imposed by p53 and retinoblastoma (Rb) tumour suppressor proteins. If replicative senescence or apoptosis are not triggered (Campisi, 2003), cells continue to divide and accumulate more dysfunctional telomeres. Then, DNA repair mechanisms (**section 1.1.4**) inappropriately repair two different unprotected ends resulting in dicentric chromosomes which are unstable and break during cell division. This process can escalate to cycles of breakage-fusion-bridge (BFB) that result in chromosomal rearrangements (**Figure 1.7**) (McClintock, 1941, Ma *et al.*, 1993). This stage is referred to as telomere-driven crisis and occurs very early during tumourigenesis. It can result in cell death (M2) caused from high levels of chromosomal instability (CIN) which are deleterious for cell viability. However, crisis provides a source of genetic variation (Murnane, 2012, Maciejowski and de Lange, 2017) that can generate tumour heterogeneity and upon selection, drive clonal evolution and malignant progression. Ultimately, reactivation of telomerase (or engaging ALT) stabilises TL and restores the protective function which allows cells to escape crisis and become immortal (Wright *et al.*, 1989, Shay *et al.*, 1991). Therefore, cells that escape crisis present a rearranged genome that provides them with tumorigenic characteristics that lead to cancer progression (Kim, 1994; Meyerson 1997) (**Figure 1.8**).

Telomere fusions of short telomeres, providing evidence of a telomere-driven crisis, have been observed in human cancers including colorectal cancer (Roger *et al.*, 2013), multiple myeloma (Hyatt *et al.*, 2017) and chronic lymphocytic leukaemia (CLL) (Lin *et al.*, 2010).

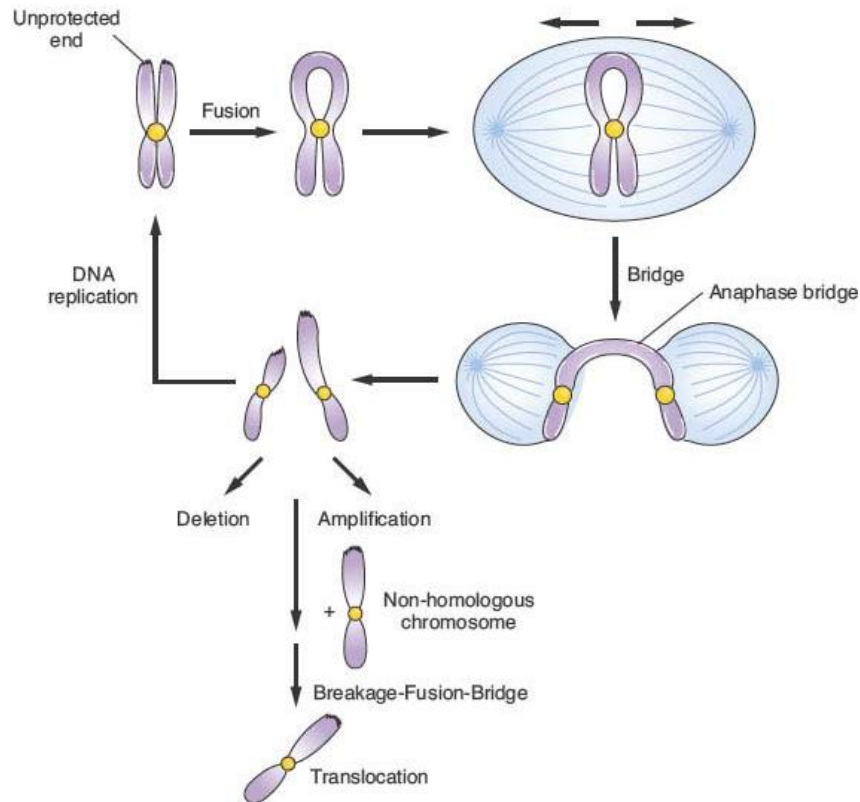


Figure 1.7. Breakage-Fusion-Bridge cycle

Unprotected telomeres can fuse, and when the cell divides each chromatid is pulled to opposite poles forming an anaphase bridge. The breakage of the bridge results in a chromatid with a deletion and another with an amplification. The chromosome ends can fuse again with a sister chromatid after DNA replication or with a non-homologous chromosome and lead to chromosomal rearrangements. Figure acquired from Oncohemakey (2016) website (URL: <https://oncohemakey.com/mechanisms-of-genomic-instability-2/>)

Cancer-causing genomic changes that have been observed in human malignancies and may occur as a consequence of dysfunctional telomeres and BFB cycles include deletions or loss of heterozygosity (LOH), amplifications, translocations and aneuploidy. In addition, kataegis (localised hypermutation), chromothripsis (chromosome shattering) and tetraploidisation (whole genome reduplication) have recently been discovered (Lo *et al.*, 2002, Sellmann *et al.*, 2016, Maciejowski and de Lange, 2017). BFB cycles can be stopped either by chromosome healing which consists of the addition of new telomeric repeats by telomerase or ALT, or by the acquisition of another chromosome end by translocation (Murnane, 2006, Zschenker *et al.*, 2009).

Even before telomeres become critically short, they can become unprotected with the loss-of-function of members of the shelterin complex and initiate telomere fusions (Denchi and de Lange, 2007). Experimental disruption of TRF2 results in telomere fusions with a mutational profile typical of C-NHEJ DNA repair mechanism (Smogorzewska *et al.*, 2002, Celli and de Lange, 2005) that is distinct from those observed in human cells with critically short telomeres (Capper *et al.*, 2007). In addition, POT1 mutations can also lead to dysfunctional telomeres that promote chromosomal instability and initiate cancer progression, particularly in CLL (Ramsay *et al.*, 2013, Gu *et al.*, 2017).

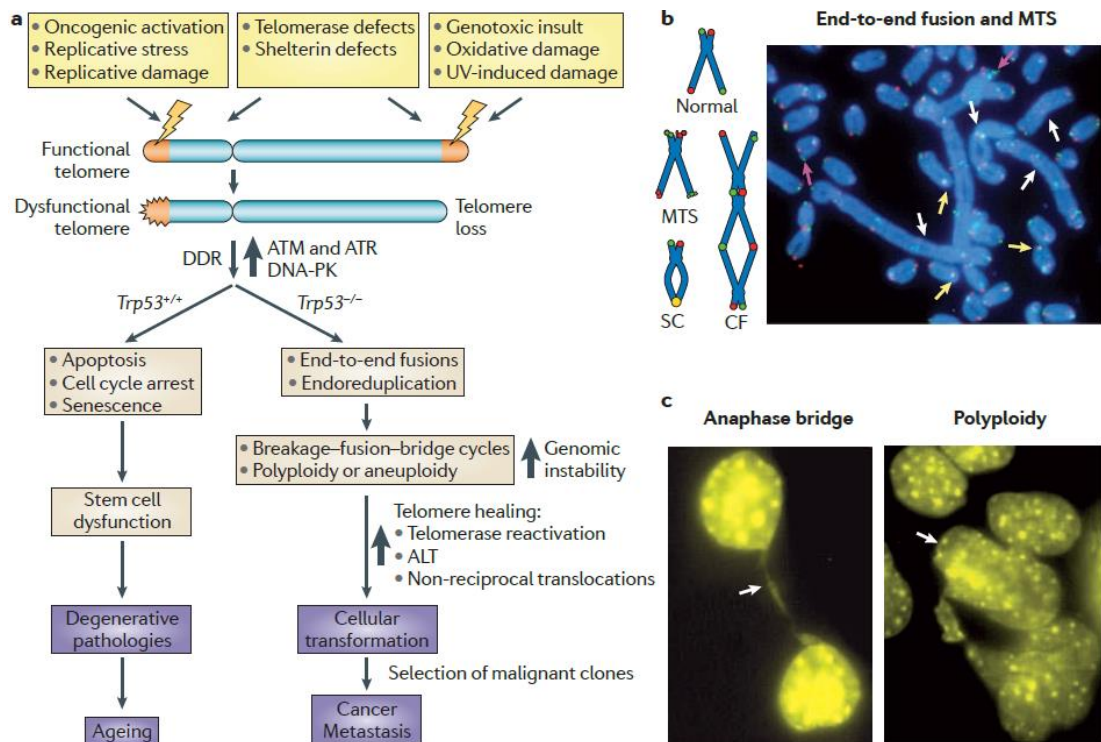


Figure 1.8. The role of telomeres in ageing and cancer.

(A) Sources of DNA damage that lead to dysfunctional telomeres or loss of telomere. Extremely short or unprotected telomeres elicit the DNA damage response (DDR) that will result in ageing or cancer depending on the p53. If the tumour suppressor p53 is intact, cell cycle arrest, replicative senescence or apoptosis will be triggered, resulting in tissue degeneration. In contrast, if p53 is inactivated, cells enter a telomere-driven crisis characterised by **(B)** end-to-end fusions and **(C)** anaphase bridges. **(C)** Polyploid cells are also observed since cells can bypass mitosis and re-enter S phase, reduplicating the genome. Cycles of BFB result in increased genomic instability that may lead to malignant transformation after telomere healing. **(B)** Chromosome fusion (CF) and concatenation indicated with a white arrow, multitelomeric signals (MTS) indicated with a purple arrow and sister chromatid fusions (SC) with a yellow arrow. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Cancer (Martinez and Blasco, 2011), copyright (2011).

1.2.2 Reactivation of telomere maintaining mechanisms in cancer

To escape crisis, cells need to reactivate telomerase. In human cancers, 85-90% of patients express telomerase while the remaining 10-15% engage ALT. This confers cells immortalisation with an indefinite replicative capacity (Kim *et al.*, 1994, Reddel *et al.*, 2001). The *hTERT* gene locus is located in Chr5p15.33. Point mutations in the *TERT* promoter, other activating mutations, amplifications of this gene or genomic rearrangements that locate *TERT* near an enhancer, have been identified in several types of cancer. Overall, this suggests that upregulation of telomerase is required for cancer progression (Beroukhim *et al.*, 2010, Huang *et al.*, 2013, Valentijn *et al.*, 2015, Heidenreich and Kumar, 2017). Recent research has provided insights into how *TERT* promoter mutations (TPM) contribute to tumorigenesis. Chiba *et al.*, (2017) propose that TPM extend cells lifespan delaying replicative senescence by stabilising the shortest telomeres. If DNA damage checkpoints are lost, when the amount of critically short telomeres increases, cells enter a long period of reduced telomere-driven genomic instability, compared to crisis. Finally, telomerase is further upregulated by unknown factors driving cell immortality (Chiba *et al.*, 2017).

TERT promoter mutations, the majority at -146 and -124 positions from the ATG start site, may generate *de novo* binding sites for transcription factors including GA-binding protein (GABP), ternary complex factor (TCF) and E twenty-six (ETS)-domain, that activate telomerase expression (Horn *et al.*, 2013, Huang *et al.*, 2013, Bell *et al.*, 2015). Human cancers with the highest frequency of *TERT* promoter mutations are glioblastoma (83%), melanoma (71%), urothelial carcinomas (66%), hepatocellular carcinomas (47%) and medulloblastomas (21%) (Lazzerini-Denchi and Sfeir, 2016).

Rearrangements involving *TERT* have been involved in telomerase upregulation in B-cell malignancies including CLL (Nagel *et al.*, 2010, Schilling *et al.*, 2013) and neuroblastoma (Peifer *et al.*, 2015, Valentijn *et al.*, 2015). Amplifications of the 5p chromosome end that included the *hTERT* locus in addition to overexpression of telomerase were observed after escape from a telomere-driven crisis in HCT116^{DN-hTERT} cells in culture (Jones *et al.*, 2014). However, it remains to be assessed whether BFB cycles or 5p sister chromatid telomere fusions, which can lead to gene amplification, could be another source for telomerase reactivation.

Furthermore, telomerase can be upregulated and promote tumorigenesis through cytokines, hormones (Kyo *et al.*, 2008), ETS transcription factors (Dwyer and Liu, 2010,

Gladych *et al.*, 2011), inactivation of tumour suppressors, activation of oncogenes (Lin and Elledge, 2003), and epigenetic modifications (Zhu *et al.*, 2010, Li *et al.*, 2011) that also regulate *TERT* transcription.

1.2.3 Telomere length and telomere dysfunction: a powerful prognostic marker

Shorter telomeres were observed in cancer cells compared with their normal tissue counterparts (Hastie *et al.*, 1990, de Lange *et al.*, 1990) and a later study of telomere fusions provided further support of a role for telomere dysfunction in cancer initiation (Ducray *et al.*, 1999). In addition, critically short telomeres have been detected in human cancers and associated with poor prognosis and reduced survival. Examples include breast cancer, chronic lymphocytic leukaemia, multiple myeloma and myelodysplastic syndrome (Simpson *et al.*, 2015, Strefford *et al.*, 2015, Hyatt *et al.*, 2017, Williams *et al.*, 2017). Therefore the measurement of TL and telomere dysfunction is a powerful prognostic tool in several malignancies. Further information about TL as a prognostic marker in CLL is provided in **section 1.3.2**. Distinct methods to measure TL and telomere dysfunction are discussed below.

1.2.3.1 Terminal Restriction Fragment (TRF) analysis

Usually described as the “gold standard”, TRF was the first technique used for the measurement of TL. This relatively cheap method is based on the digestion of genomic DNA using restriction enzymes with cleavage sites throughout the genome except for the subtelomeric and telomeric regions. Telomeres are resolved by size using agarose-gel electrophoresis and visualisation is performed by in-gel hybridisation or by southern blotting with a telomere-specific probe. The TL from all chromosomes are visualised as a smear and average TL is assessed by comparing to a DNA ladder (Allshire *et al.*, 1989, Hastie *et al.*, 1990, Kimura *et al.*, 2010).

However, TRF has some disadvantages. Large amounts of high-quality DNA are required (>1µg), TRF is not specific to individual telomeres and does not recognise individual short telomeres. Degraded DNA can provide inaccurate results, the inclusion of subtelomeric DNA overestimates the TL and only the mean TL can be obtained. TL results will vary depending on restriction enzymes used in addition to the presence of polymorphisms in subtelomeric or telomeric DNA. Still, it may be a more useful technique for long telomeres since PCR-based methods have amplicon-size limitations.

1.2.3.2 Quantitative Polymerase Chain Reaction (Q-PCR)

The principle of Q-PCR is based on the detection of fluorescence as PCR amplification occurs. Fluorescent probes are added to the PCR reaction and when they bind to the DNA amplicon, the fluorescent signal is captured. The intensity of the signal increases in proportion to the amount of DNA amplified.

A low cost high-throughput PCR-based method to measure TL was developed to overcome the need for large amounts of DNA (Cawthon, 2002). A pair of primers that target the C- and G-rich sequence and have mismatches to prevent dimerization are designed to amplify the telomeric sequence. The longer the telomere, the more binding sites are available for the fluorescent probes, which results in increased telomere amplification product (T). To quantify the TL, a single-copy gene (S) is amplified and the T/S ratio is calculated to allow comparative measurement.

To avoid inaccurate results obtained from different amount of DNA added due to pipetting error, the technique was adapted to amplify T and S in the same tube. This was called monochrome multiplex quantitative PCR (MMqPCR) (Cawthon, 2009). A later adaptation called absolute telomere length (aTLqPCR) incorporated a standard curve of known TL to obtain a mean TL (Kb) measurement instead of a T/S ratio (O'Callaghan and Fenech, 2011).

For either technique, high quality DNA is required; however, in smaller amounts (ng) compared with TRF. Some disadvantages include the requirement of a standard to obtain an absolute measurement, the TL variability among replicates, the absence of information about the TL distributions and lack of recognition of individual telomeres.

1.2.3.3 Quantitative fluorescence *in situ* hybridisation (Q-FISH)

Q-FISH examines the chromosomes within cells in metaphase or in interphase nuclei by hybridising with a fluorescent telomeric probe in conjunction with a non-specific DNA stain (such as DAPI) that allows visualisation of chromatin (Lansdorp *et al.*, 1996, Krejci and Koch, 1998). Q-FISH measures TL with a higher resolution but results are obtained in Telomere Fluorescent Units (TFUs) which can be problematic to convert into DNA length.

Individual telomeres can be identified on metaphase chromosomes, which shows the TL heterogeneity within the same cell in addition to end-to-end fusions and chromosomes lacking telomeres. However, mitotically active cells are required and, as for TRF, a minimal TL is essential for the probe to hybridise.

1.2.3.4 Flow-FISH

The combination of flow cytometry with Q-FISH resulted in flow-FISH. This technique allowed the high-throughput analysis of TL in distinct peripheral blood subpopulations (Rufer *et al.*, 1998).

1.2.3.5 Single TElomere Length Analysis (STELA)

STELA is a high-resolution long-range PCR-based assay that allows the measurement of individual telomeres at specific chromosome ends (Baird *et al.*, 2003). The method consists of a linker that is complementary to the telomeric 3' ss-overhang that is ligated at the 5' end, introducing a unique sequence that can be targeted with a specific primer. PCR is performed with an oligonucleotide that anneals to the linker, and a chromosome specific primer targeting the subtelomeric region of the same chromosome arm. STELA can be performed at the 2p, 9p, 11q, 12q, 16q, 17p, 18q and XpYp chromosome ends (Britt-Compton *et al.*, 2006). Only small amounts of high-quality genomic DNA (picogram) are required, extremely short telomeres can be detected and the sensitivity of the technique allows a very accurate measurement of the TL. The relation with TRF-TL measurements is linear although consistently less for STELA (mean 1.3Kb) (Baird *et al.*, 2003). STELA has facilitated the detailed study of telomere erosion and the end replication problem (Baird *et al.*, 2003). Despite being a very accurate technique it may not be suited for high-throughput analysis or for the study of very long telomeres (>20Kb).

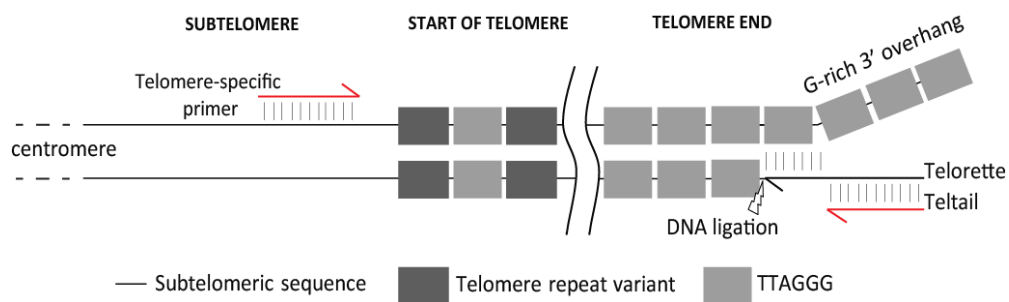


Figure 1.9. Representation of STELA PCR.

The Telorette (Tel2) linker is comprised of a 7bp-sequence at the 3' end that is complementary to the telomeric repeat sequence; as a consequence, the Tel2 anneals to the terminal 3' G-rich ss-DNA telomeric overhang. Tel2 is ligated to the 5' of the C-rich telomeric strand, through a putative ligase activity of Taq polymerase (Baird *et al.*, unpublished data). The 5' end of the Tel2 linker also contains a unique sequence identical to the Teltail primer. The Teltail primer, together with a Telomere-specific primer, are utilised to amplify the telomere in subsequent PCR cycles. Figure adapted from Baird *et al.*, 2003.

The precise measurement of TL using STELA (Baird *et al.*, 2003) facilitated the identification of a subgroup of patients with shorter telomeres and a poor outcome in several cancer types including chronic lymphocytic leukaemia, multiple myeloma, breast cancer, colorectal adenomas and other solid tumours (Lin *et al.*, 2010; Lin *et al.*, 2014; Hyatt *et al.*, 2017; Simpson *et al.*, 2015; Roger *et al.*, 2013; Letsolo *et al.*, 2017). Patient stratification was based on the TL range at which telomere fusions were detected using a PCR-based approach (Capper *et al.*, 2007).

1.2.3.6 Detecting telomere fusions

End-to-end telomere fusions can be observed by FISH in metaphase chromosomes (Ducray *et al.*, 1999) or with a telomere fusion assay (Capper *et al.*, 2007). The telomere fusion PCR consists of a combination of chromosome specific primers located about 3Kb upstream from the start of the telomere in the same orientation. In the presence of a fusion event, primers are in close spatial proximity and the fusion event can be amplified. PCR products are separated by size through agarose gel electrophoresis, followed by Southern blotting to transfer to a membrane. Hybridisation with chromosome-specific probes reveals which telomeres are involved in the fusion events and sequence content can be revealed by reamplification and purification of amplicons for Sanger sequencing (Capper *et al.*, 2007).

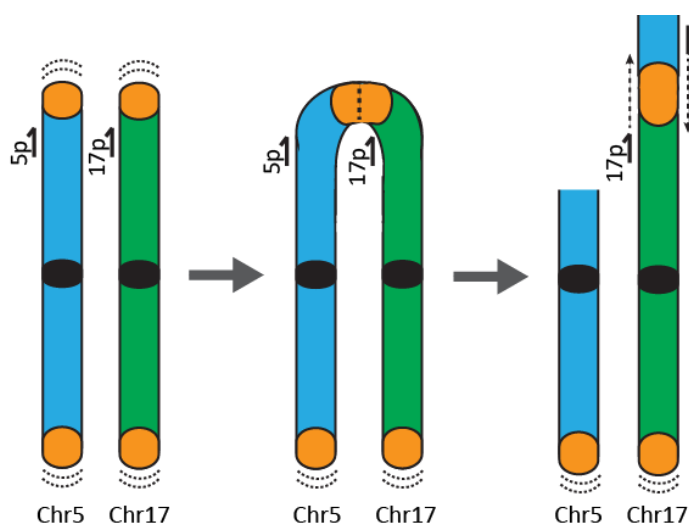


Figure 1.10. Telomere fusion PCR.

Cartoon representation of chromosomes 5 (blue) and 17 (green) with dysfunctional telomeres followed by the 5p-17p inter-chromosomal fusion and breakage. Dotted lines indicate how the 5p and 17p-specific primers amplify the fusion event.

The assay was originally performed with primers that amplified the 17p and XpYp telomeres (Capper *et al.*, 2007). Soon after, primers targeting the 21q (Lin *et al.*, 2010) and 16p families of telomeres were added (Letsolo *et al.*, 2010).

Recently, a specialised paired-end Next Generation Sequencing (NGS) of telomere fusion amplicons was developed for the high-throughput sequence analysis of telomere fusions in colorectal cell lines with induced DSBs (Liddiard *et al.*, 2016).

1.3 TELOMERE LENGTH IN CHRONIC LYMPHOCYTIC LEUKAEMIA

1.3.1 Chronic lymphocytic leukaemia

Chronic Lymphocytic Leukaemia is the most common form of adult leukaemia in Western countries, representing about one third of leukaemia cases, with an incidence rate of about 1/20000 individuals per year. Males are more susceptible than females and about 94% of cases are over 50 years old at diagnosis, with a median age of 70 (WHO, 2014, Howlader N, 2017). However, rare cases of CLL have been identified in children and teenagers which most likely relates to an inherited predisposition (Spier *et al.*, 1985, Demir *et al.*, 2014).

CLL results in a progressive accumulation of mature clonal CD19⁺ CD5⁺ B cells in bone marrow, blood and lymphoid tissues. The clinical course of the disease is highly heterogeneous. A significant proportion of patients (about 50%) remains asymptomatic and do not require treatment while others have an aggressive disease and progress quickly. The prognosis of CLL can range from months to several decades and has been a major clinical challenge (WHO, 2014, Guarini *et al.*, 2003). An indolent condition with low CLL-like cells called monoclonal B cell lymphocytosis (MBL) is a precursor to CLL. However, only 1-2% of MBL cases per year progress to CLL requiring treatment (Strati and Shanafelt, 2015). Moreover, a small proportion of cases with CLL (about 10%) can progress into Richter's syndrome, a more aggressive non-Hodgkin lymphoma (Jain and O'Brien, 2012).

1.3.1.1 Symptoms, diagnosis and staging

CLL symptoms include swollen spleen, enlarged lymph nodes, tiredness, weight loss and night sweats. Given the abnormalities in their immune system, patients can be subjected to frequent opportunistic infections (Oscier *et al.*, 2012). CLL diagnostic tests include the absolute lymphocyte count (ALC) that requires at least 5000 B cells/ μ L, immunophenotyping and/or a peripheral blood film. On occasion a lymph node biopsy may be required.

There are two different staging systems: Binet system (European) and Rai system (American) (Binet *et al.*, 1977, Rai *et al.*, 1975). They categorise patients in early (Binet A, Rai 0), intermediate (Binet B, Rai I and II) or advanced (Binet stage C, Rai III and IV) disease based on the criteria stated in **Table 1.1**.

Table 1.1. Binet and Rai classification systems.

Disease stage	Binet system	Rai system
Early	<ul style="list-style-type: none"> • Haemoglobin ≥ 10 g/dL • Platelets $\geq 100,000/\text{mm}^3$ • Enlarged lymph sites < 3 	<ul style="list-style-type: none"> • Lymphocytosis: lymphocytes in blood $> 15000/\text{mL}$, and $> 40\%$ lymphocytes in the bone marrow
Intermediate	<ul style="list-style-type: none"> • Haemoglobin ≥ 10 g/dL • Platelets $\geq 100,000/\text{mm}^3$ • Enlarged lymph sites ≥ 3 	<ul style="list-style-type: none"> • Lymphocytosis • Enlarged lymph sites • Splenomegaly or hepatomegaly or both
Advanced	<ul style="list-style-type: none"> • Haemoglobin < 10 g/dL, • Platelets $< 100,000/\text{mm}^3$ • Any number of enlarged lymph sites 	<ul style="list-style-type: none"> • Lymphocytosis • Anaemia (haemoglobin level < 11.0 g/dL or haematocrit $< 33\%$) • Thrombocytopenia (platelets $< 100,000/\text{mm}^3$)

*Binet classification system is based on enlargement of lymphoid sites (cervical, spleen, liver, axillary and inguinal lymph nodes). Rai classification system is based on the accumulation of lymphocytes at different sites: lymphocytosis (in the blood), lymphadenopathy (enlarged lymph nodes), splenomegaly (enlarged spleen) and hepatomegaly (enlarged liver).

1.3.1.2 Prognostic markers

Prognostic markers are biological characteristics that can predict patients' response to therapy or the course of the disease. For CLL, prognostic markers include the immunoglobulin heavy-chain variable region gene (*IGHV*) mutational status for which unmutated *IGHV* ($>98\%$ homology to germline) is associated with a more aggressive disease (Hamblin *et al.*, 1999). In addition, expression of CD38 antigen ($>30\%$ CD38+ CLL) and zeta-associated protein 70 ZAP-70 ($>20\%$ ZAP-70+ CLL cells) associate with worse prognosis characterised by shorter overall survival (Durig *et al.*, 2003, Malavasi *et al.*, 2011). Other prognostic markers are chromosomal aberrations including trisomy, amplification and deletion (reviewed in **section 1.4.1**), particularly 11q (*ATM*) and 17p (*TP53*) deletion which provide additional information about poor response to genotoxic treatments (Amaya-Chanaga and Rassenti, 2016, Ghamlouch *et al.*, 2017). Recently, TL has emerged as promising independent prognostic marker in CLL that is also predictive of response to treatment (reviewed in **section 1.3.2**).

1.3.2 Telomere length is a prognostic marker in CLL: 20 years of knowledge

The first evidence that TL and telomerase activity could be prognostic tools in CLL was reported from a cohort of 58 CLL patient samples using TRF almost 20 years ago. Short TL (<6Kb) inversely correlated with high telomerase activity and was associated with poor survival (Bechter *et al.*, 1998). Five years later, an association of shorter TL with unmutated *IGHV* gene status, identifying a subgroup with the worst prognosis, was established from a cohort of 61 CLL patient samples using TRF (Hultdin *et al.*, 2003). Longer TL was observed in post-germinal centre CLL cells (mutated CLL) consistent with telomerase reactivation at the germinal centre in normal B cells (Weng *et al.*, 1997, Hu *et al.*, 1997, Norrback *et al.*, 2001). In addition, Walsh *et al.*, observed that the TL of patient CLL-B cells was shorter than normal B cells of healthy individuals, indicating that CLL cells undergo extensive proliferation during disease progression (Walsh *et al.*, 2007). Although telomerase is reactivated, its expression may not be sufficient to maintain TL at all telomeres.

Measuring TL using Q-PCR and TRF in cohorts of 310 and 201 CLL patients revealed that TL was a prognostic marker and, in combination with the *IGHV* gene status, refined prediction for overall survival (OS), progression-free survival (PFS) and time to first treatment (TTFT) (Grabowski *et al.*, 2005, Ricca *et al.*, 2007). In subsequent studies, Rossi *et al.*, established a cut-off threshold for short TL at 5Kb using TRF. Their group further confirmed that TL was a robust and independent predictor of survival in a cohort of 401 CLL patients and that TL improved the accuracy of prognosis when combined with other markers (Binet stage, *IGHV* gene status and cytogenetics). Furthermore, short TL predicted progression to diffuse large B-cell lymphoma (DLBCL), a more aggressive type of leukaemia known as Richter syndrome (Rossi *et al.*, 2009).

It was 10 years after the first association of short TL with unfavourable clinical outcome that short telomeres were also linked to high-risk genomic aberrations including 11q and 17p deletion, and complex karyotype (Roos *et al.*, 2008). Roos and colleagues proposed that their results were in agreement with the “mortality stages 1 and 2 (M1 and M2) model” (Wright *et al.*, 1989, Shay *et al.*, 1991). The model states that for human cells to become immortalised they need to overcome senescence (M1) by inactivation of cell cycle checkpoints. This results in a telomere-driven crisis that initiates genomic instability and leads to apoptosis (M2). However, cancer cells escape crisis by reactivating telomerase that stabilises TL. Thus, immortalised cells are characterised by inactivation of cell-cycle

checkpoints, short but stable TL, genomic rearrangements and telomerase upregulation, as observed in the subgroup of CLL patients with a more aggressive disease.

The development of high-resolution single-molecule approaches to determine TL distributions (STELA) (Baird *et al.*, 2003) and to detect telomere fusions (Capper *et al.*, 2007, Letsolo *et al.*, 2010) has allowed our group to study telomere dysfunction in CLL patients. Telomere fusions are detected in CLL B-cells with the shortest telomeres (TL<3.81Kb) and this correlates with LOH and large scale genomic rearrangements usually observed at chromosome ends (Lin *et al.*, 2010). In addition, the mean TL (TL<2.26Kb) at which fusion can be detected (fusogenic range) is highly prognostic, particularly in early-stage CLL patients, as it identifies those patients with shorter overall survival (OS) (Lin *et al.*, 2014). Sanger sequencing of telomere fusion events amplified from CLL patient samples has revealed subtelomeric deletion, low number of TTAGGG-repeats and microhomology usage at the fusion junction (Lin *et al.*, 2010). Altogether, our group has provided evidence of telomere dysfunction and fusion of critically short telomeres during progression of CLL that is consistent with a telomere-driven crisis and which may be critical in driving genomic instability and progression of this disease (Lin *et al.*, 2010, Lin *et al.*, 2014).

Short TL has been associated with CLL driver mutations in *TP53*, *NOTCH1* and *SF3B1*, CLL high-risk genetic aberrations and more complex chromosomal alterations including copy neutral LOH, also known as uniparental disomy (UPD) (Mansouri *et al.*, 2013, Sellmann *et al.*, 2016). In particular, CLL patients with disrupted *TP53* present critically eroded TL, telomere end-to-end fusions, telomerase overexpression and increased chromosomal instability that can contribute to resistance to therapy (Gueze *et al.*, 2016).

Predicted patient survival based on the mean TL calculated using STELA (Baird *et al.*, 2003) identified that 10 years after diagnosis; patients with long telomeres (above the fusogenic range 2.21Kb) have a 91% survival rate while those with short telomeres have a 13% survival rate (Lin *et al.*, 2014). As telomere length profiles are stable over time in most cases, TL can predict CLL outcome in newly-diagnosed patients (Mansouri *et al.*, 2013, Lin *et al.*, 2014). Additionally, TL has been proved a robust prognostic marker in a cohort of 384 CLL patients from the UK LRF CLL4 clinical trial (Strefford *et al.*, 2015) and 276 from the ARCTIC and ADMIRE clinical trials (Norris *et al.*, manuscript in preparation). Therefore, measuring TL should be implemented in clinical practice. TL is an accurate prognostic marker that can help identify the subgroups of high-risk CLL patients that require earlier

treatment or that will respond poorly to standard therapies and thus require alternative treatment modalities (Pepper *et al.*, 2014).

In addition to critically short TL, telomeres can also lose their end-capping function through defective shelterin components (Poncet *et al.*, 2008, Martinez and Blasco, 2011). Faulty telosome proteins have been observed in CLL including TIN2, TPP1 and POT1 (Augereau *et al.*, 2011, Ramsay *et al.*, 2013, Ishdorj *et al.*, 2017). *POT1* is mutated in about 9% of CLL patients (Ramsay *et al.*, 2013). Most *POT1* mutations identified in CLL B-cell clones occur within the telomeric DNA-binding domain, but a small proportion of mutations are located in the TPP1-interacting domain (Ramsay *et al.*, 2013). Telomere fusions arising from dysfunctional POT1 are characterised by longer telomeres, contrasting with telomere dysfunction observed from critically short telomeres (Ramsay *et al.*, 2013, Lin *et al.*, 2010). In addition, germline mutations in POT1, TPP1 and TRF2 have been identified in familial CLL (Speedy *et al.*, 2016). A genetic predisposition to long TL caused by germline polymorphisms in telomere maintenance genes has been related to increased risk of developing CLL (Ojha *et al.*, 2016, Machiela *et al.*, 2016). The authors suggest that long telomeres may provide additional time for cells to accumulate mutations that could result in malignant transformation and telomere uncapping before reaching senescence.

Altogether it is clear that TL is a prognostic marker in CLL. Short TL is associated with telomere fusions, genome instability, a worse prognosis and potential for Richter transformation. To further assess the impact of telomere dysfunction and fusion on the CLL genome, a high-throughput characterisation of telomere fusion amplicons detected from patient CLL-B cells is required.

1.4 THE CLL CANCER GENOME

The rapid advancement in sequencing technology and computational tools has facilitated the study of patients' cancer genome, producing large quantities of data. Huge efforts have been made to translate this information into patient care, moving towards precision medicine. The study of the CLL cancer genome has incorporated genetics, epigenetics, transcriptomics, functional studies, epidemiological and clinical information to improve the knowledge of this complex and variable disease. The ultimate goal has been to improve patient stratification, monitor tumour heterogeneity, study drug resistance mechanisms and develop specialised therapeutic approaches.

1.4.1 CLL driver mutations and cytogenetic aberrations

Contributing to the International Cancer Genome Consortium (ICGC) (<http://icgc.org>), the CLL-ICGC project (2009-2014: <http://www.clongenome.es>) has characterised the cancer genome of 500 CLL patients. The study revealed a complex landscape of >2000 somatic mutations per patient. The most frequently mutated genes in CLL include *NOTCH1*, *TP53*, *POT1*, *SF3B1*, *MYD88*, *ATM*, *BRAF*, *BIRC3*, *CHD2*, *ZNF292*, *ARID1A*, *ZMYM3* and *PTPN11*. In addition, a long list of less frequent mutations that impact the pathogenesis of this leukaemia has been observed. Genes mutated in three or fewer patients include the transcription factor *IKZF3*, the oncogenes *KRAS* and *NRAS* with activating mutations, and the cell cycle regulators *CDKN1B* and *CDKN2A* with truncating mutations (Wang, 2011, Puente *et al.*, 2011, Quesada *et al.*, 2012, Martinez-Trillos *et al.*, 2013). CLL driver mutations were also identified in non-coding DNA. Mutations have been detected in the 3'UTR region of *NOTCH1* proto-oncogene which creates aberrant splicing and increases expression of this gene. In addition, mutations have been detected in the enhancer region of the B cell differentiation transcription factor *PAX5* that downregulates its expression (Puente *et al.*, 2015). The identification of driver mutations in non-coding DNA including promoters, enhancers and UTR regions highlights the importance of performing whole genome sequencing (WGS) instead of whole exome sequencing (WES) to reveal all driver mutations.

Pathways that are utilised in normal B cell function are commonly mutated in CLL. These include NOTCH signalling (lymphocyte activation); WNT signalling (regulation of proliferation); inflammatory pathways, B cell receptor (BCR) signalling and differentiation, and NF- κ B pathway (development and function); RNA and ribosomal processing; chromatin

modification; DNA damage response, cell cycle control and apoptosis (**Figure 1.11**) (Puente *et al.*, 2015, Kipps *et al.*, 2017, Ghamlouch *et al.*, 2017).

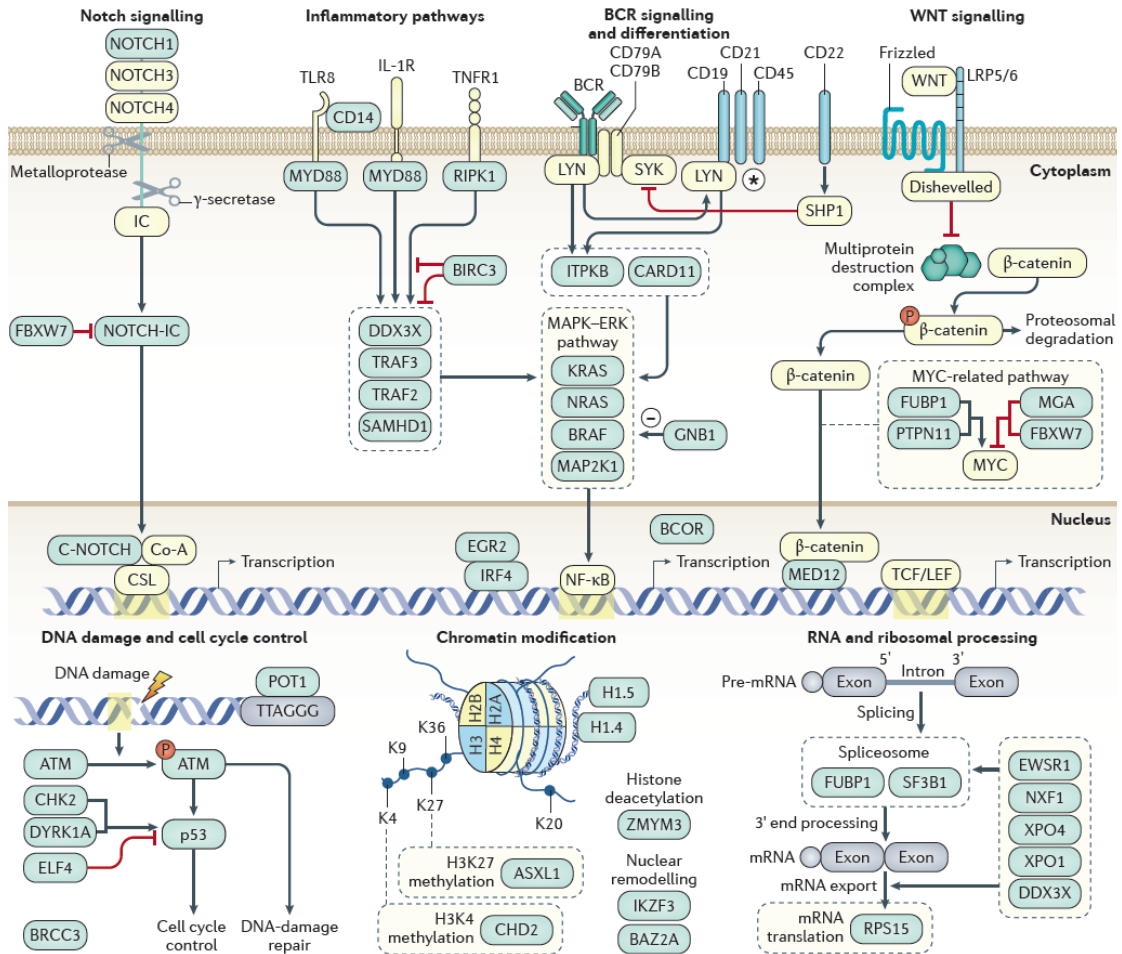


Figure 1.11. Mutated pathways in CLL.

Genes that have somatic mutation in CLL are highlighted in blue boxes. The main altered pathways in CLL include NOTCH signalling, inflammatory pathways, BCR signalling and differentiation, WNT signalling, DNA damage and cell cycle control, chromatin modification, RNA and ribosomal processing. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Disease Primers (Kipps *et al.*, 2017), copyright (2017).

Genomic aberrations can be driver if they have an impact on malignant progression by providing a selective advantage or passenger if they are random events with no impact on the disease (Stratton *et al.*, 2009, Garraway and Lander, 2013). Mutations can be inherited (germline mutations) or acquired through lifetime (somatic mutations) (Futreal *et al.*, 2004, Alexandrov *et al.*, 2015, Ju *et al.*, 2017). Specific patterns or mutational signatures have helped reveal the mutational processes causing somatic DNA damage. These include cell division and ageing, abnormal DNA editing and repair mechanisms, UV radiation, chemical carcinogens such as tobacco and fungal toxins, amongst other (Alexandrov *et al.*, 2013a).

In collaboration with Alexandrov and colleagues, the analysis of somatic mutations in CLL identified 3 distinct mutational signatures that correspond to ageing, APOBEC-induced cytosine deamination (C>T) (AID) and somatic hypermutation at the germinal centre (Alexandrov *et al.*, 2013a). Somatic hypermutation status of the *IGHV* gene has generally been used as the main prognostic factor in CLL, with patients with unmutated *IGHV* displaying a worse prognosis than patients with mutated *IGHV*. Recently, WGS of 46 CLL patients has revealed a mutational signature for each subtype: mutated *IGHV* associated with anomalous AID activity and non-coding mutations while unmutated *IGHV* related to mutations in coding DNA and an ageing signature (Burns *et al.*, 2017).

A variety of chromosomal abnormalities associated with different prognostic outcomes is found in CLL. The most common include: deletion 13q14 (50% of patients), trisomy 12 (20%), 2p gain (5-28%), 11q22-23 deletion (6-20%) and 17p13 deletion (5-10%). Other less frequent Copy Number Variations (CNV) include 8q24.21 amplification (5%), 8p deletion (5%), 15q15.1 deletion (4%), deletion of 2q37, 3p21 and 10q24 (2% each). Genomic aberrations including key genes disrupted, pathways associated and the impact on patient's prognosis is listed in **Table 1.2** and recently reviewed in Ghamlouch *et al.*, 2017.

Table 1.2. Recurrent copy number changes in CLL.

Abnormalities	Frequency (%)	Associated genes	Cellular processes affected by the alteration	Prognostic significance
del(13q14)	50	<i>MIR15A/MIR16-1, DLEU2, RB1, DLEU7</i>	Regulation of BCL2 expression, cell cycle control, NF-κB signalling	Good
trisomy 12	20	Unknown	Unknown	Good/intermediate
del(11q22-23)	6–20	<i>ATM, BIRC3</i>	DNA repair, NF-κB signalling	Poor
del(17p13)	5–10	<i>TP53</i>	Loss of tumour suppressor	Poor
del(6q21)	5–7	<i>ZNF292</i>	Transcriptional regulation	Unknown
Gain 2p	5–28	<i>XPO1, REL, BCL11A, MYCN</i>	RNA processing, NF-κB signalling, Proliferation	Poor
amp(8q24.21)	5	<i>MYC</i>	Proliferation, apoptosis	Poor
del(8p)	5	<i>TRAIL-R</i>	Apoptosis	Poor
del(15q15.1)	4	<i>MGA</i>	Transcriptional regulation	None
del(2q37)	2	<i>SP140/SP110</i>	Transcriptional regulation	None
del(3p21)	2	<i>SMARCC1/SETD2</i>	RNA splicing and DNA repair	Poor
del(10q24)	2	<i>NF-κB2</i>	NF-κB signalling	Unknown

Table adapted by permission from John Wiley and Sons: British Journal of Haematology (Ghamlouch *et al.*, 2017), copyright (2017).

Recurrent translocations have been observed in CLL patients, including chromosome 13q with distinct partners resulting in the common 13q14 deletion present in about 50% of CLL patients (Gardiner *et al.*, 1997, Struski *et al.*, 2007, Hrubá *et al.*, 2012). Furthermore, complex genomic instability including chromothripsis has been identified in several CLL patients (Stephens *et al.*, 2011, Pei *et al.*, 2012, Bassaganyas *et al.*, 2013, Salaverria *et al.*, 2015). In another study, chromothripsis in CLL was associated with shorter telomeres and it was proposed to arise from cycles of telomere BFB (Ernst *et al.*, 2016).

In addition, approximately 30 Single Nucleotide Polymorphisms (SNPs) identified from genome-wide association studies (GWAS) have been associated with CLL risk, suggesting a potential involvement in CLL pathogenesis. Most SNPs have been implicated in apoptosis, DNA damage and chromosomal stability, B-cell development and immunity. Susceptible loci related to telomeres and telomerase include *BCL2L11* at the ancestral telomere in 2q13, *TERT* in 5p15.33 and *POT1* at 7q31.33 (Cerhan and Slager, 2015, Berndt *et al.*, 2016, Kandaswamy *et al.*, 2016, Law *et al.*, 2017).

1.4.2 Tumour heterogeneity

CLL presents intra- and inter-patient tumour heterogeneity, with a variety of genomic aberrations contained in distinct clones (Crossen *et al.*, 1993, Landau *et al.*, 2013, Landau *et al.*, 2014). Chromosomal instability and somatic mutagenesis are a source of genetic variability that together with selective pressure, contribute to tumour evolution. Some events may arise early during disease progression and are therefore present in all cells (clonal), while other abnormalities arise during disease progression and are only present in a proportion of cells (subclonal). Distinct CLL subclones can be found in equilibrium within the patient; however, therapy and lymph node microenvironment can change the tumour composition and select for resistant subclones, resulting in patient relapse (Puente and Lopez-Otin, 2013, Ojha *et al.*, 2015, Del Giudice *et al.*, 2016).

The prognostic impact of CLL driver mutations and cytogenetic aberrations can vary depending on whether the mutation is a clonal or subclonal event (Nadeu *et al.*, 2016, Yi *et al.*, 2017). In addition, it has been identified that distinct subclones with mutated and unmutated *IGHV* can coexist within the same patient. It also raises questions about the cell of origin and suggests that a B-cell progenitor, before undergoing somatic hypermutation, could be the leukaemia-initiating cell (Kriangkum *et al.*, 2015, Stamatopoulos *et al.*, 2017). Therefore, assessing and characterising tumour heterogeneity is essential for both an accurate prognosis and treatment.

1.4.3 Other 'Omics' studies

Complementary omics studies have revealed alterations in the epigenome and transcriptome of CLL patients.

Epigenetic alterations modify gene expression without altering the DNA sequence, of which DNA methylation is the most common (Rodriguez-Paredes and Esteller, 2011, Dawson and Kouzarides, 2012, Feinberg *et al.*, 2016). Mutations in chromatin remodelling protein coding genes have been identified at low frequency in CLL (Puente *et al.*, 2015). Epigenomics revealed DNA hypomethylation in components important for B cell differentiation or activity and discovered 3 epigenetic signatures that describe distinct CLL subgroups: naïve or memory B-cell-like CLL for pre-germinal or post-germinal centre B-cells, respectively and an intermediate state. The classification system based on the 3 epigenetic signatures identifies the cell of origin and has proven to predict prognosis more accurately than *IGHV* status (Kulis *et al.*, 2012, Queiros *et al.*, 2015).

Transcriptomic alterations have been identified in protein-coding genes, non-coding RNAs, and pseudogenes. A different transcriptomic pattern has been observed for CLL cells compared to normal B cells including upregulation of genes involved in metabolic pathways, B-cell receptor and JAK-STAT signalling, and downregulation of genes associated to the spliceosome, ribosome and proteasome. Furthermore, two distinct CLL groups have been identified based on transcriptomal complexity, with the MAPK/ERK signalling pathway the most differentially expressed. B-cell receptor stimulation in the lymph node microenvironment has been suggested as the origin of one subgroup (Ferreira *et al.*, 2014).

1.4.4 Therapeutics and precision medicine

For those patients that require treatment, there are different options available. Compared with normal cells, cancer cells have a rapid proliferation rate and have defects in DNA repair mechanisms. Thus, chemotherapeutic drugs aim to damage the DNA to increase the genetic instability in the cell through the accumulation of DSBs that would induce cell death via apoptosis, based on a threshold hypothesis (Fu *et al.*, 2012). Chemotherapeutic drugs include alkylating agents like Chlorambucil and Cyclophosphamide, and purine analogues like Fludarabine, Cladribine and Pentostatin (Pettitt, 2003). Other treatment options target elements of the B cell receptor signalling cascade, blocking proliferation and survival. These include the Bruton tyrosine kinase (BTK) inhibitor, Ibrutinib (Byrd *et al.*, 2015), or the PI3K δ inhibitor Idelalisib (Brown *et al.*, 2014).

Advancements in cancer immunotherapy have provided a better understanding of the tumour interaction with the immune system. A therapeutic approach includes using monoclonal antibodies that target the CD20 (Rituximab) and CD52 (Alemtuzumab) antigens on the B cells surface, flagging the cells for destruction. In addition, somatic mutations have the potential to generate neoantigens, cancer-specific peptides which can be used for vaccine development to stimulate the immune system to kill tumour cells (Rajasagi *et al.*, 2014, Liu and Mardis, 2017). A promising therapeutic strategy is genetically engineered autologous T cells, known as chimeric antigen receptor (CAR) T cells, that are infiltrated into the patient to kill CD19 cells (Freeman and Gribben, 2016, Turtle *et al.*, 2017).

The primary therapeutic option in CLL, known as the gold standard, is chemo-immunotherapy combining Fluradabine, Chlorabucil and Rituximab (FCR) that improves patient survival compared with single agent therapy or FC (Robak *et al.*, 2010). Additional radiotherapy to the spleen may be done for a small proportion of cases. Bone marrow transplant or haematopoietic stem cell transplantation (HSCT) is rare given the high-risk of the procedure, the age of diagnosis and the unclear benefit compared with chemotherapy (Gribben, 2009).

Altogether, this highlights the importance of understanding the CLL genome in this new era of precision medicine. Its study has allowed a better patient stratification, a comprehensive understanding of the disease and the identification of new therapeutic targets, which in combination with functional analysis and animal studies translates into benefit for patients.

1.5 HYPOTHESIS AND AIMS OF THE PROJECT

The principal aim of this Ph.D. thesis was to investigate the impact of telomere dysfunction and fusion on the cancer genome of patients with Chronic Lymphocytic Leukaemia (CLL). With particular focus on the 5p telomere, the aim was to determine whether telomere dysfunction at this chromosome end, harbouring the *hTERT* locus, could initiate genomic instability that might result in reactivation of telomerase in cancer. The specific aims were as follows:

- To adapt STELA and the TVR assay to measure the 5p telomere length and the TVR content and compare it with the well-characterised 17p and XpYp telomeres in 57 CLL patient samples.
- To develop the telomere fusion assay at the 5p chromosome end to examine whether the 5p telomere is dysfunctional and fusogenic in CLL.
- To detect telomere fusions in a cohort of 276 CLL patient samples with short telomeres (TL<3.81Kb) and to investigate whether the frequency of fusions provides further prognostic information in this subgroup.
- To characterise telomere fusion amplicons, particularly those involving 5p, from CLL patients with the highest frequency of fusions using high-throughput paired-end Illumina sequencing.
- To investigate the impact of telomere dysfunction on the CLL cancer genome by studying which areas of the genome become incorporated into telomere fusions and whether there is an association with CLL or other oncogenic pathways.
- To determine the existence of intra-tumour heterogeneity in a CLL patient sample with a bimodal telomere length distribution. Additionally, to examine the patient's whole genome to investigate the cause of the increased fusion frequency and the patient's indolent disease.

CHAPTER 2:

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals and reagents

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

2.1.2 Laboratory materials and equipment

Plastic and glass laboratory equipment used was obtained from different sources which include Becton Dickinson, Eppendorf, Gilson, Starstedt and Thermo Scientific.

Specialised equipment for use in experiments was obtained from different sources: autoMACS Pro Separator (Miltenyi), ARIA III FACS sorter flow cytometer (Becton Dickenson), Centrifuges (MSE), QuantiFluor fluorometer (Promega), PCR thermocycler (Bio-RAD and Thermo Scientific), Electrophoresis System (Amersham), UV-Transilluminator (EPS), Heating Systems (Jencons, Grant Instruments), Hybridisation ovens (Thermo Scientific), Typhoon FLA 9500 biomolecular imager (GE Healthcare).

2.1.3 Oligonucleotides

Primers were designed based on the GRCh37/hg19 human reference with the help of the online browser Primer3 (Rozen and Skaletsky, 2000) (<http://primer3.ut.ee/>) and according to the following criteria: 24bp long (20-24bp), T_m of 60°C (59-64°C) and a 55 % GC content (45-70%). For enhanced primer annealing the sequence finished on A/T+C/G at the 3' end. Once the list of potential primers was obtained, the oligonucleotide sequences were aligned with the human genome using BLAST alignment tools (Altschul *et al.*, 1990) from Ensembl (http://grch37.ensembl.org/Homo_sapiens/Tools/Blast) and NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to exclude those with unspecific binding. In addition, *in silico*-PCR (USCS) determined the specificity of the primers

(<https://genome.ucsc.edu/cgi-bin/hgPcr>). The primers were synthesised by Eurofins Genomics. Primers that were used during this project are listed in **Supplementary Table 1**.

2.1.4 CLL patient samples

Peripheral blood samples from 36 CLL patients undergoing treatment at the University Hospital of Wales (UHW) were obtained by Professor Chris Fegan. Patient peripheral blood mononuclear cells (PBMC) from 39 CLL treatment naïve patients were obtained from the UK LRF CLL4 trial (Oscier *et al.*, 2010) and 204 previously untreated CLL patients from both the ARCTIC and ADMIRE trials (Howard *et al.*, 2017, Munir *et al.*, 2017) were obtained from the UK CLL Trials Biobank, University of Liverpool. Ethical approval and informed consent was obtained in all cases according with the ethical approval established for this study (Table 2.1).

Table 2.1. Ethical approval for each CLL cohort.

COHORT	ETHICAL APPROVAL
UHW	South East Wales Local Research Ethics Committee (LREC# 02/4806)
LRF CLL4	UK Leukaemia Research Fund (LRF) CLL4 trial UK Multicentre Research Ethics Committee (MREC). International Standard Randomised Controlled Trial (NCT 58585610).
ARCTIC	Attenuated dose Rituximab with ChemoTherapy In Chronic lymphocytic leukaemia (ARCTIC) trial. National Research Ethics Service Leeds (East) Research Ethics Committee (REC) (reference 09/H1306/54). International Standard Randomized Controlled Trial (ISRCTN16544962).
ADMIRE	Does the ADDition of Mitoxantrone Improve Response to FCR chemotherapy in patients with CLL? (ADMIRE) trial. Leeds West Research Ethics Committee and the Medicines and Healthcare products Regulatory Agency. International Standard Randomized Controlled Trial (ISRCTN42165735).

2.2 METHODS

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

2.2.1 Isolation of PBMCs

CLL blood samples from the UHW were collected in tubes containing 7.2mg K2-Ethylenediaminetetraacetic acid (K2-EDTA). Lymphocytes were isolated from whole blood by density gradient centrifugation using Ficoll-Histopaque-1077 (Sigma Aldrich). The blood was carefully layered onto the Histopaque-1077 to a ratio 1:1 and centrifuged at 840xg without the centrifuge break for 20min at room temperature. Following centrifugation, the opaque interface which contained the PMBC was carefully aspirated using a Pasteur pipette. The cells were washed by adding 10mL PBS followed by a further centrifugation step at 470xg for 10min with the brake on. The supernatant was discarded and the cell pellet was resuspended with 1-3mL (depending on the size of the cell pellet) of 1x Red Blood Cell (RBC) Lysis Buffer (10x, BioLegend) to lyse remaining red blood cells and incubated 10min in the dark at room temperature. Cells were washed with 10mL of PBS followed by a further centrifugation step at 470xg for 5min with the brake on. Cells were resuspended with PBS and counted using the Vi-Cell XR (Beckman Coulter) outlined in **section 2.2.3**.

2.2.2 Cell culture

Human Embryonic Kidney (HEK) 293 cells were cultured in 75cm² flasks containing Dulbecco's Modified Eagle Medium (DMEM, Life Technologies) in a 37°C, 5% CO₂ incubator. DMEM was supplemented with 10% v/v Fetal Calf Serum (FCS, Thermo Fisher Scientific), 2mM L-glutamine (Sigma), 100Units/mL Penicillin (Sigma) and 0.1mg/mL Streptomycin (Sigma).

HEK293 cells were passaged when cell density reached 80-95% confluency, usually after 72h. The media was removed and cells were washed with 1x phosphate-buffered saline (PBS). Then, to detach the adherent cell line from the flask, 2mL of 1x trypsin (Thermo Fisher Scientific), a proteolytic enzyme, was added to the cells and incubated at 37°C for 5min. To inactivate trypsin, cells were washed with 10mL of DMEM and a sample was taken for cell count (**section 2.2.3**). Cells were harvested and centrifuged at 470xg for 5min. Cell

pellets were resuspended to a concentration of 1×10^6 cells/mL and split between new flasks with 20mL of fresh supplemented DMEM media pre-warmed at 37°C.

2.2.3 Cell count and viability

Cells were diluted 1:10 in PBS in a cell counting cup (Beckman Coulter) and loaded into a Vi-Cell XR (Beckman Coulter). Cell counts and viability were assessed by trypan blue exclusion.

2.2.4 Cryopreservation of cells and thawing procedure

Cells ($1-10 \times 10^5$) were pelleted by centrifugation before resuspending in 1mL of cryopreservation media containing 50% of appropriate media (cell type-dependent), 25% DMEM, 15% FCS and 10% dimethyl sulfoxide (DMSO). Cells were stored in cryovials and placed in a cryofreezing container at -80°C for 24h. To ensure the samples froze at a controlled rate (1°C/min), the container was filled with isopropyl alcohol. Afterwards, cells were transferred to liquid nitrogen for long-term storage.

Once cells were removed from liquid nitrogen they were rapidly thawed in a water bath at 37°C and transferred to a new tube where 9mL of the appropriate media was added dropwise. Cells were centrifuged at 470xg for 5min, the supernatant was removed and cells were resuspended in the appropriate growth media.

2.2.5 Cell enrichment using magnetic microbeads

Isolation of CD19⁺ CLL-B lymphocytes from a patient PBMC sample was performed through positive selection with anti-human CD19 microbeads (Miltenyi) using the autoMACS Pro Separator. PBMC were washed in PBS and centrifuged for 5min at 470xg. The supernatant was discarded and the cell pellet was resuspended with 20μL/ 10^7 cells of CD19 microbeads and 80μL/ 10^7 cells of cold (4°C) magnetic-activated cell sorting (MACS) buffer (1x PBS, BSA 500mg/mL, EDTA 20mM, Miltenyi Biotech). Following 15min incubation at 4°C under rotation using a MACSmix (Miltenyi), cells were washed in 2mL/ 10^7 cells of cold MACS buffer and centrifuged for 5min at 470xg. The supernatant was discarded and cells were resuspended in 500μL cold MACS buffer ready for magnetic enrichment of CD19⁺ CLL-B lymphocytes using the autoMACS Pro Separator.

To analyse the purity of the sample, two aliquots containing 1×10^5 cells were washed with 1mL of PBS and centrifuged at 470xg for 5min. The supernatant was discarded and 2μL of human CD19-APC antibody (Mouse IgG1, clone HIB19, Biolegend) was added to an aliquot

(stained fraction) that was incubated at 4°C in the dark for 15min. After a final wash with PBS, the cells were centrifuged and the subsequent cell pellet was resuspended in 100µL of PBS. CD19⁺ stained and unstained samples were analysed using Accuri C6 flow cytometer and integrated software (BD Biosciences) assisted by Dr Lauren Elston.

2.2.6 Cell sorting by flow cytometry

Isolation of CD19⁺ CLL-B lymphocytes and CD3⁺ T lymphocytes from a CLL patient sample for WGS was achieved using the ARIA III FACS sorter flow cytometer (BD). PBMC cell pellets were resuspended in 300µL of PBS and were incubated with 5µL of the monoclonal antibodies anti-human CD19-APC (Mouse IgG1, clone HIB19, Biolegend) and anti-human CD3-FITC (Mouse IgG1, clone UCHT1, Biolegend) at 4°C for 15min in the dark. Following washing and resuspending the cells in fluorescence-activated cell sorting (FACS) buffer (PBS with 1% Fetal Bovine Serum (FBS)) to a concentration of 2×10^7 cells/mL in a 15mL falcon tube, cells were sorted using the ARIA III FACS sorter (BD) into 15mL falcon tube containing 2ml cold FBS. Purity of the samples was analysed before and after sorting and data analysis was performed using FlowJo (Tree star) by Dr Catherine Naseriyan and Professor Chris Pepper.

2.2.7 DNA extraction: Phenol/Chloroform/Isoamyl Alcohol

Genomic DNA from UHW and LRF CLL4 trial CLL patient samples containing > 80% B cells or from ARCTIC and ADMIRE trials CLL patient samples was extracted from pellets of 3×10^6 cells by standard RNase A, Proteinase K, phenol/chloroform extraction (Sambrook, 1989).

Cells were lysed for 15h at 45°C with 300µL lysis buffer (10mM Tris-HCl pH8, 100mM NaCl, 0.5% Sodium Dodecyl Sulphate (SDS) and 5mM EDTA pH8), 30µg RNase A (10mg/mL; Sigma) and 60µg proteinase K (20mg/mL; Sigma). Following cell lysis, 300µL of phenol/chloroform/isoamyl alcohol (25:24:1, pH8, Sigma) was added to the cell lysate and rotated for 30min at room temperature using a tube rotator. To separate the phases, the mix was centrifuged for 5min at 16,000xg. The aqueous phase and interphase were transferred into a new 1.5 mL tube containing 300µL phenol/chloroform/isoamyl alcohol and the mix was rotated for a further 20min. Following centrifugation at 16,000xg for 5min, the aqueous phase containing the DNA was transferred to a fresh 1.5mL tube. To precipitate the DNA, 30µL (1/10 volume) of sodium acetate (3M, pH5.3, Sigma) and 900µL of 100% ice-cold ethanol (-20°C) were added to the mix. This was incubated at -20°C for 15h for optimal precipitation of the DNA.

Following centrifugation (5min 16,000xg), the DNA pellet was washed with 70% ice-cold ethanol. Once all the residual ethanol was completely evaporated by air drying, the DNA was resuspended in Tris-HCl buffer (10mM, pH8).

2.2.8 DNA quantification

DNA concentration was quantified using the Fluorescent DNA Quantitation Kit (BioRad) in triplicate for each sample. 1xTEN assay buffer (10x stock; 100mM Tris, 10mM EDTA, 2M NaCl, pH7.4) solution was prepared with double-distilled ddH₂O. Then, Hoechst 33258 DNA intercalator fluorescent dye (1mg/mL stock) was added to a final concentration of 0.1µg/mL. To calibrate the fluorometer, 2mL of 1xTEN solution containing Hoechst was read (blank). For the 500ng standard, 10µL of calf thymus DNA (100µg/mL) was added to a clean cuvette with 2 mL 1xTEN solution containing Hoechst. Sample DNA was thawed at 37°C to ensure a homogeneous solution and 1µL of DNA with 2 mL 1xTEN solution containing Hoechst was measured. If concentration was close to the standard range, DNA was further diluted with Tris-HCl (10mM, pH8) and measured again. The mean of the reads for each sample was calculated.

2.2.9 Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) technique was used to amplify specific DNA fragments. Depending on the application an optimised protocol was used as described in the following sections.

2.2.9.1 Conventional PCR

For a conventional PCR, typically 50ng of DNA per μL were used for each reaction and the reagents were added as listed in **Table 2.2**. The composition of the Taq buffer (10x) used consisted of 75mM Tris-HCl (pH8.8), 20mM $(\text{NH}_4)_2\text{SO}_4$ and 0.01% Tween-20 unless otherwise stated. The concentration of the Taq polymerase was 500U/100 μL .

Table 2.2. Conventional PCR reaction mix

Component [stock]	1x reaction (μL)	[1x reaction]
ddMilliQ H ₂ O	14.05	
Taq buffer (10x)	2.5	1x
MgCl ₂ (25mM)	2	2mM
dNTPs (100mM)	0.3	1.2mM
forward primer (10 μM)	2.5	1 μM
reverse primer (10 μM)	2.5	1 μM
Genomic DNA (50ng/ μL)	1	50ng
Taq polymerase (500U/100 μL)	0.25	1.25U
TOTAL	25 μL	

The reactions were cycled using a Tetrad thermal cycler (Bio-Rad) under the conditions stated in **Table 2.3**. On occasion, for optimisation of the primer annealing temperature, a gradient (52-68°C) was used.

Table 2.3. Conventional PCR cycling conditions

Denaturation	Annealing	Elongation	Short-term storage
94°C (20s)	52-68°C (30s)	68°C (1 min/Kb)	10°C
32 cycles			∞

2.2.9.2 STELA PCR

STELA PCR was used to amplify individual telomeres from specific chromosome ends as depicted in **Figure 1.9**. DNA extractions were diluted to a final concentration of 250pg/ μ L (1.25ng/ μ L DNA for 5p STELA*) in a volume of 40 μ L of Tris-HCl (10mM, pH7.5) containing 250pM of Telorette2 (Tel2) linker. 1 μ L of the DNA/Tel2 mixture was added to a 10 μ L reaction as listed in **Table 2.4**.

Table 2.4. STELA PCR reaction mix

Component [stock]	1x reaction (μ L)	[1x reaction]
ddMilliQ H ₂ O	4.98	
Taq buffer 10x	1	1x
MgCl ₂ (25mM)	0.8	2mM
dNTPs (100mM)	0.12	1.2mM
Telomere specific primer (5 μ M)	1	0.5 μ M
Teltail primer (5 μ M)	1	0.5 μ M
DNA/Tel2 mix	1	250pg*/250pM
Taq/PWO (10:1U)	0.1	0.5/0.05U
TOTAL	10	
*Or 1.25ng for 5p STELA		

For each sample, 6 PCR replica reactions were set up all samples received the same master mix (**Supplementary Figure 1**). Reactions were cycled using a Tetrad thermal cycler (Bio-Rad) as indicated in **Table 2.5**.

Table 2.5. STELA PCR cycling conditions

Denaturation	Annealing	Elongation	Short-term storage
94°C (20s)	59-65°C (30s)	68°C (8min)	10°C
22 cycles			∞

Annealing temperature varied depending on the telomere-specific primer used (59°C for 5p and 17p, 65°C for XpYp STELA). The number of PCR cycles was decreased compared with a conventional PCR to limit the amount of amplicon produced. This limited the size of the resulting band, improving resolution during telomere length (TL) quantification (**section 2.2.12**).

For each experiment, a positive control for STELA PCR was included. The PCR product of a previous sample known to work would be loaded in the gel together with the new PCR products for agarose gel electrophoresis (**section 2.2.10.1**) followed by Southern blotting (**section 2.2.11**).

2.2.9.3 TVR PCR

TVR PCR was used to measure the content of specific telomere variant repeats (TVRs). For each sample, a total of three reactions per chromosome end was set up, one for each combination of telomere-specific primer and TVR primers (TAG-TelW for TTAGGG, TAG-TelX for TGAGGG or TAG-TelY for TCAGGG) according to **Table 2.6**. It results in the amplification of several fragments of distinct size that followed by gel electrophoresis (**section 2.2.10.1**) and Southern blotting (**section 2.2.11**) indicate the position of the canonical repeat and the specific TVRs targeted.

Table 2.6. TVR PCR reaction mix

Component [stock]	1x reaction (μL)	[1x reaction]
ddMilliQ H ₂ O	1.58	
Taq buffer 10x	1	1x
MgCl ₂ (25mM)	1.2	3mM
dNTPs (100mM)	0.12	1.2mM
Telomere specific primer (5μM)	2	1μM
TVR primer (5μM)	2	1μM
DNA (50ng/μL)	2	100ng
Taq polymerase (10U)	0.1	0.5/0.05U
TOTAL	10	

The reactions were cycled using a Tetrad thermal cycler (Bio-Rad) under the conditions annotated in **Table 2.7**.

Table 2.7. TVR PCR cycling conditions

Denaturation	Annealing	Elongation	Short-term storage
96°C (20s)	67°C (30s)	70°C (3min)	10°C
19 cycles			∞

Representation of TVR PCR is depicted in **Figure 3.6A**.

2.2.9.4 Telomere fusion PCR

The telomere fusion PCR contains 5 primers that detect 24 distinct chromosome ends: 3 chromosome specific (5p, 17p and XpYp) and 2 that target subtelomeric families (16p and 21q). The assay was used to amplify telomere fusion amplicons as illustrated in **Figure 1.10**.

10 PCR reactions were set up for each sample as stated in **Table 2.8** and protocol depicted in **Supplementary Figure 2**.

Table 2.8. 5p8:17p6:XpYpM:16p1:21q1 fusion PCR reaction mix

Component [stock]	1x reaction (μ L)	[1x reaction]
ddMilliQ H ₂ O	3.48	
Taq buffer 10x	1	1x
MgCl ₂ (25mM)	0.8	2mM
dNTPs (100mM)	0.12	1.2mM
5p8 (10 μ M)	0.5	0.5 μ M
17p6 (10 μ M)	0.5	0.5 μ M
XpYpM (10 μ M)	0.5	0.5 μ M
16p1 (10 μ M)	0.5	0.5 μ M
21q1 (10 μ M)	0.5	0.5 μ M
DNA (50ng/ μ L)	2	100ng
Taq/PWO (10:1U)	0.1	0.5/0.05U
TOTAL	10	

Reactions were cycled using a Tetrad thermal cycler (Bio-Rad) under the conditions listed in **Table 2.9**.

Table 2.9. Fusion PCR cycling conditions

Denaturation	Annealing	Elongation	Short-term storage
94°C (20s)	62°C (30s)	68°C (8min)	10°C
25 cycles			∞

2.2.9.5 Fusion reamplification PCR

To reamplify a fusion event for subsequent sequence characterisation, the PCR reactions were diluted 1:20 in double-distilled water. A 30 μ L PCR reaction was set up containing 3 μ L of diluted fusion PCR, the combination of forward and reverse primers needed to reamplify the fusion event, together with the reagents stated in **Table 2.10**. Primers were selected based on the chromosome ends involved in the fusion event.

Table 2.10. Fusion reamplification PCR reaction mix

Component [stock]	1x reaction (μL)	[1x reaction]
ddMilliQ H₂O	17.92	
Taq buffer 10x	3	1x
MgCl₂ (25mM)	2.4	2mM
dNTPs (100mM)	0.36	1.2mM
Forward primer (10μM)	1.5	0.5mM
Reverse primer (10μM)	1.5	0.5mM
DNA (1:20)	3	
Taq/PWO (10:1U)	0.3	0.5/0.05U
TOTAL	30	

The reactions were cycled using a Tetrad thermal cycler (Bio-Rad) under the same conditions as for the fusion PCR (**Table 2.9**).

2.2.10 Gel electrophoresis

2.2.10.1 Gel electrophoresis for STELA, TVR and fusion PCR products

The STELA, TVR and fusion PCR amplicons were resolved by a low density 0.5% agarose Tris-acetate-EDTA gel electrophoresis causing the negatively charged, long DNA fragments to migrate towards a positive electrode and therefore allow separation by size. The agarose (Roche) was dissolved in boiling Tris-acetate-EDTA (40mM Tris base, 20mM acetic acid, 1mM EDTA) and was cooled down to 50°C. Then, 20µL Ethidium Bromide (EtBr) (0.625mg/mL) was added to intercalate with the DNA and allow its visualisation under the UV light.

PCR products contained 1x Ficoll based loading dye (5% bromophenol blue, 5% xylene, 15% Ficoll) to make the sample more dense and allow it to sink on the well, as well as to determine the location of the smaller molecules that would run quicker in the gel. For STELA and TVR, the samples were loaded in individual pre-set wells and were subsequently resolved on a 40cm long gel at 120V for 16h. For a suitable resolution in the fusion analysis a 20cm long gel was sufficient. Therefore the 40cm long gel was used and a second row of pre-set wells was set up for space optimisation. DNA fragments were resolved by electrophoresis at either 50V for 16h or 200V for 3h. For enhanced resolution, the gel was submerged in 1xTAE and incubated at 4°C by circulating it through a cooling system.

2.2.10.2 Standard PCR products

PCR products were resolved on 0.7-1% agarose (Geneflow) Tris-acetate-EDTA gel electrophoresis on a smaller 10-20cm gel at 100V for 2h. The density of the agarose gel would vary depending on the expected size of the DNA fragment; a higher concentration would be used for smaller fragments.

2.2.10.3 Visualisation of PCR products

DNA bands stained with the fluorescent dye EtBr were visualised using a UV-transilluminator (EPS). The UV-light (wavelength of 210-285nm) excites the EtBr that emits orange light at 605nm. For STELA, TVR and the fusion assay, EtBr was used to visualise the DNA ladders on the gel to therefore cut the appropriate length in preparation for Southern blotting (**section 2.2.11**).

2.2.11 Southern blotting

Southern blotting is a technique that allows detecting specific DNA sequences. The steps required include: transfer of the DNA fragments from an agarose gel to a nylon membrane, radiolabelling with a specific DNA probe and visualisation.

2.2.11.1 DNA transfer from an agarose gel to a nylon membrane

The STELA, TVR and fusion gels were washed twice (6min/wash) in depurination buffer (0.25M HCl) that removed purine bases breaking the DNA in smaller fragments to facilitate the transfer. The gel was subsequently rinsed and washed for 15min in denaturation buffer (1.5M NaCl, 0.5M NaOH) that denatured the double-stranded DNA to enable hybridisation. A positively charged nylon membrane (Hybond-XL, GE Healthcare) was placed on top of the agarose gel followed by a stack of paper towels and applied weight. The negatively-charged single-stranded DNA fragments were transferred by alkaline Southern blotting during 4h using the same denaturation buffer that passed through the gel and filtered to the paper towels.

2.2.11.2 Synthesis of the radiolabelled DNA probe

25ng of probe DNA in a total volume of 45µL TE buffer (10mM Tris-HCl and 1mM EDTA) was heated to 96°C for 5min to denature the DNA. The mixture was incubated on ice for 5min, the probe was added to Rediprime Labelling system (GE Healthcare), together with 4µL of [³³P]dCTP (3000Ci/mmol), and was incubated in a water bath at 37°C for 1h. The radiolabelled probe was generated by incorporating α-³³P-dCTP into the synthesised DNA. Afterwards, 1µL of radiolabelled DNA ladder (1kb and 2.5kb) was added together with 50µL of ddMilliQ water to stop the reaction. Radiolabelled probes were stored in the fridge at 4°C up to 2 weeks.

STELA and TVR products were detected using a probe specific for the telomere repeat sequence. Telomere-specific probes were generated by PCR using pairs of telomere-adjacent primers specific for each of the chromosomes ends involved in the fusion assay.

2.2.11.3 Hybridisation

The hybond membranes (GE Healthcare) were incubated in 15mL church buffer (0.5M Sodium phosphate dibasic (Na₂HPO₄), 7% SDS, 1% Bovine Serum Albumin (BSA), 1mM

EDTA) at 60°C in the hybridisation oven for 30min to reduce background hybridisation signal.

The radiolabelled probe was denatured by heating at 95°C for 5min. Hybridisation was initiated by injecting 25µL of denatured radiolabelled probe into the hybridisation bottle (Thermo Scientific) which was rotated for 15h overnight at 60°C. These conditions were used for high stringency hybridisation to avoid unspecific binding of the probe.

2.2.11.4 Removing unbound probe

To remove excess unbound probe or bound non-specifically to the membrane, blots were washed three times with a wash buffer (0.1x sodium chloride sodium citrate (SSC), 0.1% SDS) and incubated at 60°C for 30min for a longer wash which was subsequently repeated. Stringent washes at high temperature and low salt concentration allowed the stability of perfectly matched hybrids only. STELA or TVR blots were transferred onto filter paper and dried at 60°C for 25min. For fusion blots, excess water was removed by placing the membranes onto paper towels. Then, the blots were subsequently wrapped in cling film ready for exposure.

2.2.11.5 Visualisation of radiolabelled blots

The radiolabelled southern blots were left for 15h exposure in a cassette with a molecular dynamic phosphorimager screen (Amersham) on top. The screen was scanned and the hybridised fragments were detected by phosphorimaging with a Typhoon FLA 9500 biomolecular imager (GE Healthcare) and visualised using Molecular dynamics ImageQuant TL (GE Healthcare).

The radiolabelled probe was stripped from fusion blots with 500mL of boiling 0.1% SDS for 1h that was repeated three times prior to hybridising with the next probe.

Sequential hybridisations in the order 5p – 17p – XpYp – 16p – 21q were undertaken with each of the telomere-specific probes. Each probe was stripped using boiling 0.2% SDS.

2.2.12 Gel analysis and statistics

For STELA blots, the molecular weights of the resolved telomeres were determined using the TotalLab TL120 imaging software (Nonlinear Dynamics) based on comparison with the ladder run on the gel detected with the ladder probe. The measurements were exported onto Microsoft Excel where the mean and standard deviations for each sample were calculated once the distance between the binding site of the telomere adjacent primer to the start of the telomere was subtracted. Statistical analysis from the data obtained from STELA and TVR was performed using GraphPad Prism 5.

2.2.13 Sanger Sequencing

2.2.13.1 DNA extraction from agarose gel

For DNA Sanger sequencing of a specific genomic location or a telomere fusion amplicon, PCR products were electrophoresed on an agarose gel (detailed in **section 2.2.10**) and the DNA band was excised using a scalpel. The DNA was extracted from the excised agarose band using the High Pure PCR product purification kit (Roche).

The DNA was eluted in 35µL of elution buffer (10mM Tris-HCl, pH8.5) and was stored at -20°C for subsequent use.

2.2.13.2 Sequencing reaction

A 10µL PCR reaction was set up containing purified DNA, sequencing primer and BigDye Terminator Cycle sequencing mix (v3.1) (Applied Biosystems) according to **Table 2.11**.

Table 2.11. Sequencing PCR reaction mix

Component [stock]	1x reaction (µL)	[1x reaction]
Purified DNA	4.4	
Primer (1µM)	1.6	0.16µM
BigDye reagent	4	1X
TOTAL	10	

The reactions were cycled Tetrad thermal cycler (Bio-Rad) under the following conditions:

Table 2.12. Sequencing PCR cycling conditions

Denaturation	Annealing	Elongation	Short-term storage
96°C (10s)	50°C (5s)	60°C (4min)	10°C
25 cycles			∞

Following sequencing, 10µL ddMilliQ H₂O was added to the product and the reactions were purified using Dye Ex 2.0 spin kit (Qiagen) to remove unincorporated dye terminators by gel-filtration. The cleaned product was sequenced by the Central Biotechnology Service (CBS), Cardiff University.

2.2.13.3 Analysis of sequences

The sequenced DNA was viewed using Sequence Scanner v1.0 and analysed using Basic Local Alignment Search Tool (BLAST) software on the Ensembl and NCBI websites.

2.2.14 Paired-end Next Generation Sequencing

2.2.14.1 Pre-sequencing PCR clean-up

Before sequencing, PCR reaction products were purified of PCR reagents (primers, nucleotides, enzymes and salts) by magnetic separation using Agencourt AMPure XP purification kit (Beckman Coulter) along with the DynaMag-2 magnet (Life technologies).

For each sample, 200 fusion PCR reactions were prepared and 4 random PCR reactions were taken for quality control validation by southern blotting before the remaining reactions were pooled for purification. Following fusion PCR validation, the reactions from each sample were pooled and a 5µL pre-cleared aliquot was taken. The remaining product was split equally in 1.5mL tubes and gently mixed with AMPure XP reagent at room temperature (1.8mL AMPure XP/1mL PCR product). Following a 5min incubation period at room temperature the PCR products were bound to the paramagnetic beads. Since the paramagnetic beads were coated with carboxyl molecules, in the presence of polyethylene glycol (PEG) and salt, the beads reversibly bound the negatively charged DNA (>100bp). The tubes were transferred to the magnet and incubated for 2min to allow the separation of the beads and PCR products from contaminants (PCR reagents). The supernatant was subsequently discarded.

The beads and PCR products were washed twice with 70% ethanol to remove contaminants. The ethanol was discarded and the beads were left to air-dry for 20min to remove the residual ethanol. Tubes were removed from the magnet and the PCR products were eluted from the beads with nuclease-free water (volume (V); $V_{\text{elution}} = \frac{1}{4} \times V_{\text{original}}$). Tubes were returned to the magnet for 2min to allow the beads to separate from the solution containing the eluted DNA sample. The post-clean PCR products were pooled and a 5 μ L aliquot was taken.

To verify the functionality of the purification of the PCR reactions the pre- and post-purification aliquots were run on a 0.5% agarose gel, southern blotted and hybridised with the fusion radiolabelled probes.

2.2.14.2 Illumina HiSeq4000 PE100 paired-end NGS

To sequence telomere fusion amplicons, paired-end sequencing was chosen. As the distance between each paired-read is known, it allows the detection of gene fusions, chromosomal aberrations and novel transcripts, or in this case, telomere fusions (**Figure 2.1**). In addition, it allows sequencing repetitive regions with a high mapping quality.

Illumina HiSeq4000 100PE paired-end NGS of the telomere fusion amplicons from 1 CLL patient sample was performed by Oxford Genomic Centre (UK) and the remaining 8 CLL patient samples by BGI Tech (Hong Kong).

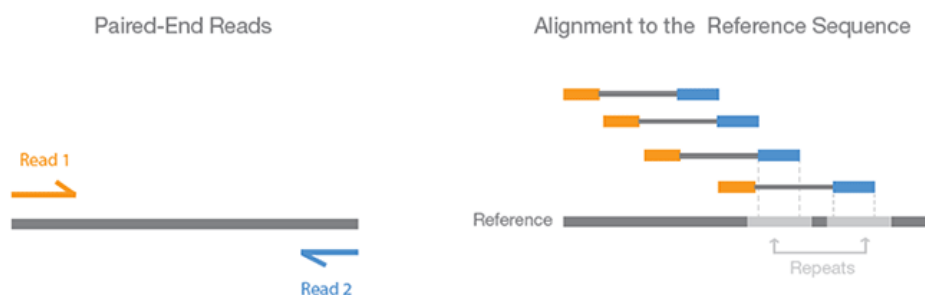


Figure 2.1. Paired-end sequencing and alignment.

Taken from Illumina, Inc. (www.illumina.com). Paired-end NGS, sequences both ends of a DNA fragment resulting in improved alignment of the reads, particularly in repetitive regions of the genome and facilitates the detection of telomere fusions.

2.2.14.3 Data processing and quality controls (QC) with bioinformatics tools

The bioinformatics approach used was based on a novel pipeline developed by Dr Kevin Ashelford (Liddiard *et al.*, 2016). Dr Kez Cleal performed the data handling, QC and mapping strategies. The downstream analysis including manual curation and investigation of telomere fusions were performed by myself.

Raw sequencing data, containing the DNA sequence and the quality scores, were downloaded from the server in the text-based format FASTQ files. Read quality was measured from the FASTQ files using the FastQC (v0.11.2) quality control software package (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Mapping success and insert size-range were also calculated from the data. The read yield, amount of DNA sequenced per sample, were in the range (10-15Mb) of data.

Reads were trimmed to remove sequence tags and primers using Trimmomatic (version 0.30) (Bolger *et al.*, 2014) and data for each sample was processed for observation of two different types of telomere fusion events: intra- and inter-chromosomal. The mapping approaches taken for each case are explained.

For analysis of **intra-chromosomal** or sister chromatid telomere fusion events, discordant read-pairs mapping to a single subtelomere (5p, 17p or XpYp) were filtered on orientation and only discordant paired-reads both in forward (F) orientation representing head-to-head fusions and not genomic input were selected for analysis.

For the **inter-chromosomal** analysis, reads were mapped to the unmasked human genome build 37 (hg19) reference sequence containing appended subtelomeric sequences (Stong *et al.*, 2014) and the 5p, 17p, XpYp, 16p family and 21q family subtelomeric sequences (based on HCT116 human cell line reference). From the list of discordant read-pairs, those with an unmapped read/mate or a read/mate in the reverse strand were excluded.

Genome alignment using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009) software produced Sequence Alignment Map (SAM) files that were then converted into the compressed binary version (BAM files), together with their index (BAI files). BAM files were then sorted, merged and filtered according to the intra- and inter-chromosomal analysis using SAMtools (Li *et al.*, 2009). Spreadsheets containing the results obtained from the intra- and inter-chromosomal analysis and files for visualisation of the data were generated for manual curation and downstream analysis.

The customised scripts used for the mapping and analysis (available at GitHub https://github.com/nestornotabilis/GenomeResearch_2016_scripts) consisted of a Bash wrapper script that embedded Perl scripts. To process some of the scripts, an associated Java code was needed <https://github.com/nestomotabilis/WGP-Toolkit>.

2.2.14.4 Manual curation and downstream analysis

To exclude ambiguous or poor quality mappings, discordant read-pairs were filtered on MAPQ>0. The read coverage on its own was not a good indicator of authenticity due to the proportion of these rare and unique telomere fusion events. Therefore, each of the telomere fusions identified from the read-pair mapping was further validated manually. Information considered included the mapping quality score for paired-reads, the orientation of the reads and the proportion of softclipping that may identify a sequence aligning elsewhere in the genome. The events were visualised using the tool Integrative Genome Viewer (IGV) (Robinson *et al.*, 2011) and reconstructed and interrogated using BLAST against the human genome and the subtelomeric sequences. Telomere fusions were classified into distinct type of events, microhomology at the fusion junction was examined and subtelomeric deletion was measured at sister chromatid fusion events. In addition, the signature of telomere fusions for each CLL patient was investigated and genomic loci incorporated into fusions were studied in detail.

A summary of the workflow for the analysis of the telomere fusion amplicons is depicted in **Figure 2.2**. More detailed information about the manual curation and downstream analysis can be found in results chapter 4.

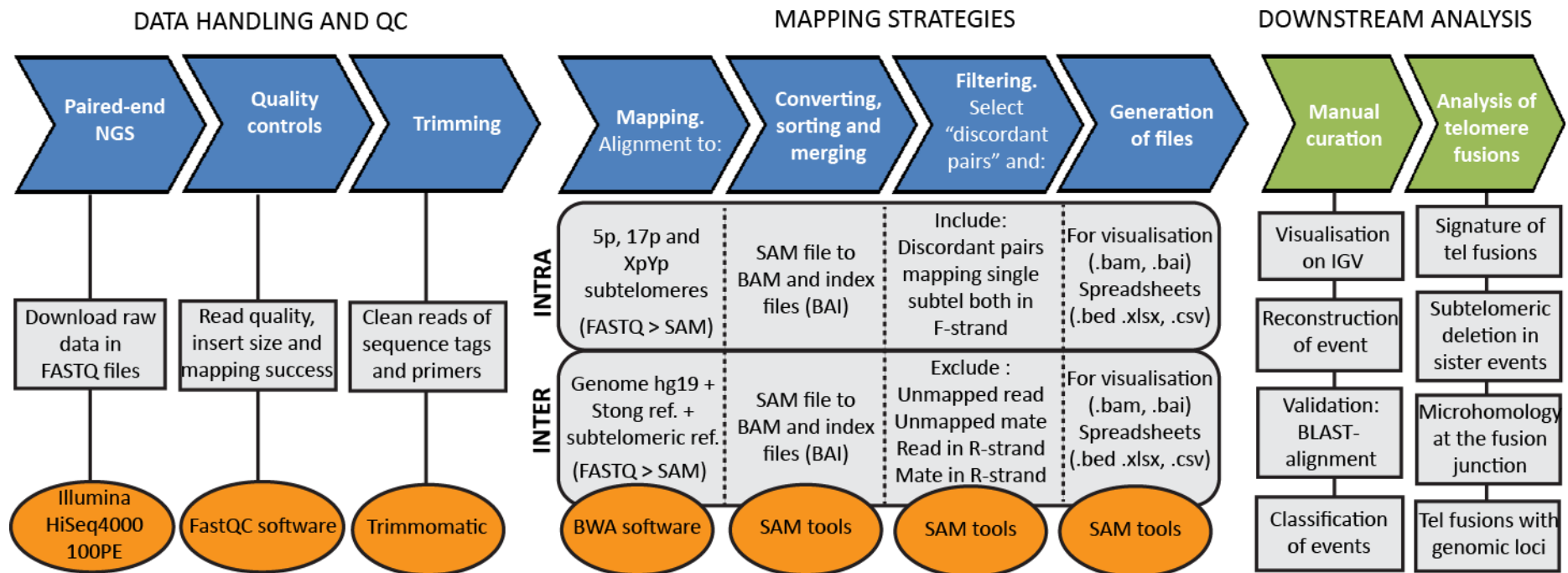


Figure 2.2. Workflow for the NGS analysis of telomere fusion amplicons.

Pipeline for the detection of telomere fusion events using bioinformatics tools, followed by manual curation and downstream analysis for each CLL patient sample. Sequencing data of telomere fusion amplicons from 9 CLL patient samples were obtained from BGI Tech after they performed HiSeq4000 PE NGS. Data handling and QC were performed following intra- and inter-chromosomal mapping strategies. Finally, manual curation and downstream analysis were performed. Stong reference is the human subtelomeric sequence mapped by Stong *et al.*, 2014.

2.2.15 Whole Genome Sequencing

2.2.15.1 Illumina HiSeq2000

CLL patient genomic DNA samples derived from tumour CD19⁺ B cells (30µg) and control-matched CD3⁺ T cells (2µg) were sent to BGI Tech Solutions (Hong Kong) for Illumina HiSeq2000 PE whole genome sequencing (WGS) with 60x and 30x coverage, respectively. Library preparation including fragmentation and adapter ligation, DNA amplification resulting in cluster generation and sequencing by synthesis, were performed at BGI tech.

DNA mapping, QC and variant calling were performed by Dr Kez Cleal. Manual curation and downstream analysis were performed by myself.

2.2.15.2 Data processing and quality controls (QC) with bioinformatics tools

Once the raw data (FASTQ files) were obtained from the Illumina pipeline, primers, tags and low quality reads were removed using Trimmomatic (version 0.30 (Bolger *et al.*, 2014)). Clean data was aligned to the human genome build 37 (hg19) reference sequence using the BWA software (Li and Durbin, 2009). The SAM file obtained was then converted to BAM file and its index BAI files using SAMtools (Li *et al.*, 2009).

Quality checks of reads were performed before and after alignment to the reference genome.

2.2.15.3 Variant calling: SNVs and copy numbers

Variant discovery analysis from the tumour sample compared to the normal, was performed using MuTect (Cibulskis *et al.*, 2013) and Somatic Snipper (Larson *et al.*, 2012) for calling single nucleotide variants (SNVs) and cn.MOPS (Klambauer *et al.*, 2012) for discovering copy number variations (CNVs).

2.2.15.4 Manual curation and downstream analysis

Selected variants identified were manually validated by visualisation in IGV and Sanger Sequencing. Downstream analysis included investigating clonality within the sample, the *IGHV* gene mutational status, the impact of mutations and the presence of other forms of genomic instability such as translocations, uniparental disomy, kataegis and chromothripsis within the tumour sample. More information can be found in results chapter 5.

CHAPTER 3:

CHARACTERISATION OF THE 5P TELOMERE: MEASUREMENT OF THE TELOMERE LENGTH AND INVESTIGATION OF TELOMERE DYSFUNCTION AND FUSION.

3.1 ABSTRACT

The catalytic subunit of telomerase (*hTERT*) is one of the most distal genes on 5p. Translocations involving this locus have previously been detected in CLL and may impact on telomerase activity and disease progression. In this study it was examined whether dysfunction at the 5p telomere could lead to chromosomal instability that affects telomerase expression.

In this chapter, high-resolution single-molecule approaches were adapted to precisely measure telomere length (TL) and to detect unique fusion events at the 5p telomere. Data were consistent with the 5p telomere displaying similar TL profiles to other unique telomeres, including 17p and XpYp. Rare telomere fusions involving 5p were identified in 22.6% (40/177) of CLL patient samples with fusions, which may provide insights into the mechanism of *hTERT* translocation in cancer. Particularly interesting are sister chromatid fusions that could lead to *hTERT* amplification and inter-chromosomal fusions with 17p, that could result in loss of the tumour suppressor gene *TP53*.

To further investigate telomere-driven chromosomal rearrangements in CLL, 276 samples from patients with short telomeres (TL<3.81 Kb) were screened. Telomere fusions were detected in 198 samples (71.7%). Different frequencies of fusions were observed amongst CLL patients with short telomeres; however, it did not correlate with TL. The fusion frequency failed to provide further prognostic information, although 15 months shorter progression free survival (PFS) and shorter overall survival (OS) was observed among CLL patients with the highest frequency of fusions than patients with no fusions detected. Moreover, 15 samples with the highest fusion frequency ($>4.20 \times 10^{-5}$ per diploid genome) were identified and 9 were selected for high-resolution NGS characterisation of telomere fusions in the CLL genome, featured in Chapter 4.

3.2 INTRODUCTION

Extensive telomere erosion or unprotected telomeres result in the loss of telomere function and the induction of senescence (Harley *et al.*, 1990, Bodnar *et al.*, 1998). If cell cycle checkpoints are not intact, cells can continue to divide and enter a telomere-driven crisis characterised by telomere fusions and the initiation of breakage-fusion-bridge (BFB) cycles that can result in large-scale genomic rearrangements (McClintock, 1941, Ma *et al.*, 1993). The stabilisation of telomere length (TL) is required for cells to escape crisis and gain replicative immortality, this is considered to be an essential step for malignant progression (Kim *et al.*, 1994, Meyerson *et al.*, 1997). Telomerase is reactivated in 85% of human malignancies (Kim *et al.*, 1994, Meyerson *et al.*, 1997). The Telomerase Reverse Transcriptase (*hTERT*) is the catalytic subunit of the telomerase holoenzyme that mediates the addition of hexameric repeats to extend telomeres and increase the lifespan of the cell (Greider and Blackburn, 1985, Morin, 1989). In humans, *hTERT* is located proximal to the 5p telomere (Chr5p13.33). In cancer, upregulation of telomerase can be caused by *hTERT* promoter point mutations that create new binding sites for transcription factors, translocations of the gene near an enhancer or amplifications of the *hTERT* locus (Bell *et al.*, 2015, Horn *et al.*, 2013, Huang *et al.*, 2013, Valentijn *et al.*, 2015). Rearrangements of this locus have been observed in neuroblastoma and associated with more aggressive disease (Valentijn *et al.*, 2015, Peifer *et al.*, 2015). In B-cell malignancies including CLL, rearrangements including *hTERT* have been involved in telomerase upregulation (Nagel, 2010; Schilling *et al.*, 2013). Genomic rearrangements of the 5p chromosome were observed in 3/8 CLL patients consistent with a chromothripsis-like pattern, including gains of the *hTERT* locus (Salaverria *et al.*, 2015). In addition, amplification of 5p, including the *hTERT* locus, coincident with telomerase upregulation was observed in cells after an escape from a telomere-driven crisis in culture in HCT116 cells transfected with a dominant-negative *hTERT* (DN-*hTERT*) construct (Jones *et al.*, 2014).

Telomere fusions have been detected in patient CLL-B cells with short telomeres (TL<3.81Kb) which correlated with loss of heterozygosity (LOH) and increased genomic instability (Lin *et al.*, 2010). Moreover, CLL patients within the mean TL range at which fusions are detected (TL<2.26Kb), present a more aggressive disease with shorter survival (Lin *et al.*, 2014).

As sister chromatid fusion events have the power of gene amplification and given the proximity of *hTERT* to the 5p telomere, in this study it is hypothesised that 5p-5p intra-chromosomal telomere fusions have the potential to initiate amplification at the *hTERT* locus and that this may be important in the progression of CLL.

3.3 AIMS OF THE PROJECT

The purpose of this chapter was to assess whether the 5p telomere is dysfunctional and can undergo fusion in CLL patients, particularly sister chromatid fusions that can lead to gene amplification. This was achieved by adapting STELA and the single-molecule telomere fusion assay to characterise the 5p telomere.

The aims of this chapter were as follows:

- To adapt STELA and the TVR assay to measure the 5p telomere length and the content of telomere variant repeats (TVRs).
- To compare the 5p telomere length with the well-characterised 17p and XpYp telomeres in order to evaluate whether the 5p telomere exhibits differential dynamics that may drive dysfunction and fusion at this telomere.
- To adapt the single-molecule telomere fusion assay to include the 5p telomere
- To investigate whether the 5p telomere undergoes dysfunction and fusion in a cohort of 276 CLL patients with short telomeres (TL<3.81Kb).
- To identify CLL patients with the highest frequency of telomere fusions.
- To assess whether the frequency of fusions could further stratify CLL patients into prognostic subgroups.
- To investigate changes in 5p TL and LOH during a telomere-driven crisis in culture.

3.4 RESULTS

3.4.1 5p STELA measures the telomere length of the 5p chromosome arm in human cell lines and CLL patients

A pair of primers was designed with the 367bp adjacent to the start of the human chromosome 5p telomere repeat array (**Supplementary Table 2**) and they were tested in gradient PCR using Human Embryonic Kidney (HEK) 293 human cell lines. 5p specificity was confirmed with subsequent Sanger sequencing of the amplicons (**Figure 3.1**).

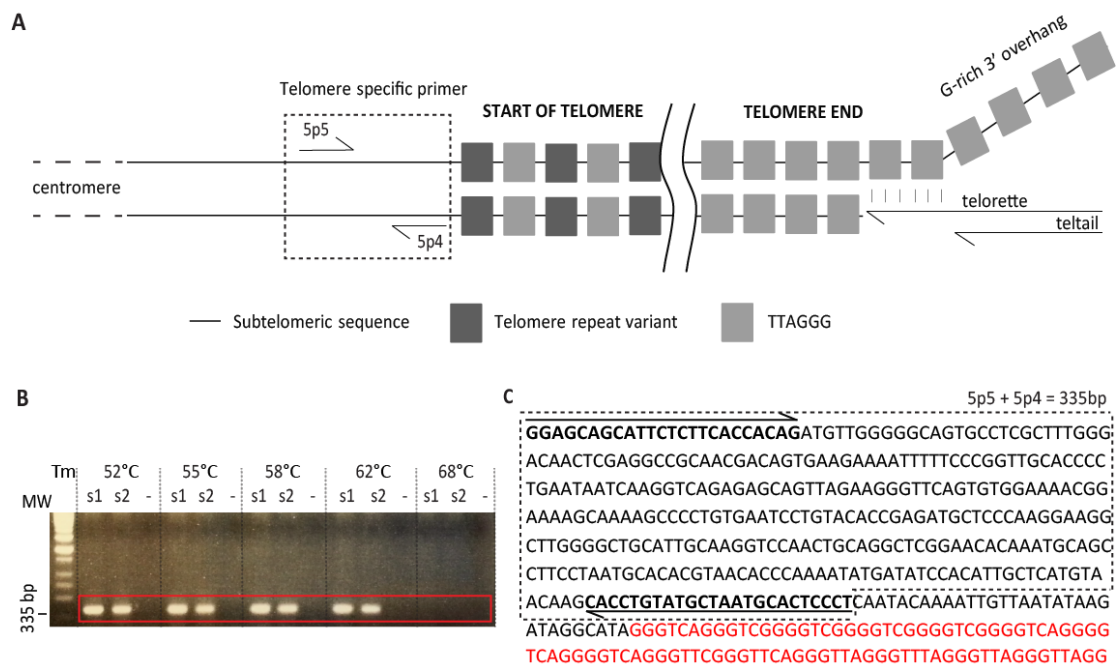


Figure 3.1. Development of 5p STELA.

(A) Adaptation of Baird *et al.*, 2003 showing the localisation of STELA oligonucleotide primers and linkers at the 5p chromosome end. (B) Amplicons generated with a gradient PCR with the 5p4 and 5p5 primers for s1: HEK293 cells, s2: CLL patient sample, -: ddH₂O (negative control). For further experiments T_m 59°C was selected for 5p STELA. (C) Sequence of the product obtained with the 5p4 and 5p5 primers (335bp) used as a 5pA telomere adjacent specific probe in Southern blots highlighted in the dashed box. 5p primers indicated with bold and underlined. Start of the telomere indicated with red.

To validate the specificity of the 5p STELA and confirm reproducibility, the 5p TL was determined in two clonal populations from the HCT1080 and MRC5 cell lines for which XpYp TL profiles were previously characterised. A single homogeneous XpYp TL distribution was observed for HCT1080 clone 5, since all cells derive from the same clone they present

similar TL. In contrast, a bimodal distribution was observed from the MRC5 clone 13 population indicative of two XpYp alleles of discrete and different lengths (**Figure 3.2A**).

If the 5p telomere presented distinct allelic TLs or 5p STELA amplified telomeres of different chromosome ends, more than one TL distributions would be observed. However, a single 5p TL distribution was observed for both HT1080 and MRC5 clonal populations, indicating that 5p STELA was specific for the 5p chromosome end only and not other chromosome ends (**Figure 3.2A**).

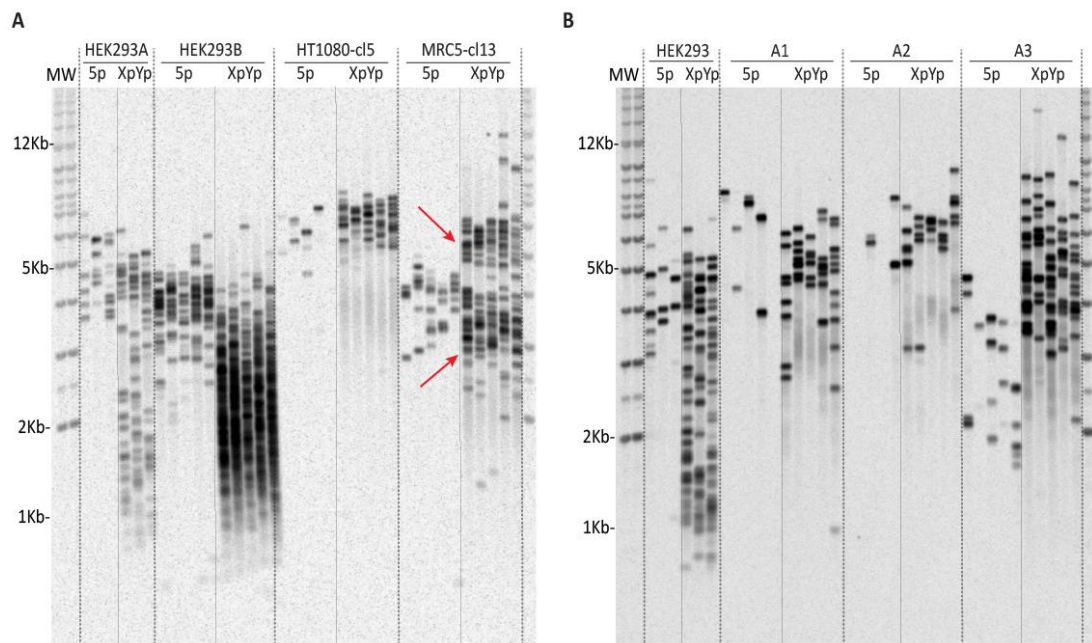


Figure 3.2. Validation of the specificity and optimisation of 5p STELA.

(A) Comparison of 5p and XpYp STELA for the cell lines HEK293, HT1080 clone 5 and MRC5 clone 13. The two different alleles of MRC5-cl13 marked by red arrows. (B) 5p and XpYp STELA for 3 different CLL patient samples. STELA products were generated using 250pg DNA per reaction with the following primers: 5p5 and XpYpE2, and detected with the telomere probe. Each single band represents a telomere.

The accurate assessment of 5p TL in cell lines and CLL patient samples was hampered by the lower numbers of amplicons compared to XpYp STELA using the same template DNA (250pg) (**Figure 3.2**). To increase the number of amplifiable molecules detected with 5p STELA, the DNA template input was adjusted by titration until comparable numbers of amplified molecules to the XpYp and 17p STELA was achieved, allowing proper statistical comparisons of TL in patient samples. For 5p STELA, 5 times more DNA than for 17p and XpYp STELA was required (**Figure 3.3**).

To provide an additional confirmation of the assay specificity and to detect short 5p telomeres that may not hybridise with the telomere-repeat probe, a probe (5pA) that hybridises to the 5p telomere-adjacent DNA sequence was designed using the 5p4 and 5p5 primers (**Figure 3.1C**).

A comparison of 5p STELA on CLL patient samples detected with both the 5pA probe (**Figure 3.3A**) and the telomere probe (**Figure 3.3B**) enabled the detection of three different types of telomeres depending on the pattern of hybridisation. Type I included most of the telomeres, which were detected with both the telomere and the 5pA probes. Type II comprised very short telomeres (TL<1Kb) detected with the telomere adjacent probe but not hybridised with the telomere probe because of the lack of TTAGGG-repeats. Lastly, type III consisted of telomeres (TL>1Kb) detected with the telomere probe but not the 5pA DNA adjacent probe indicating they might not contain enough telomere-adjacent DNA sequence to hybridise.

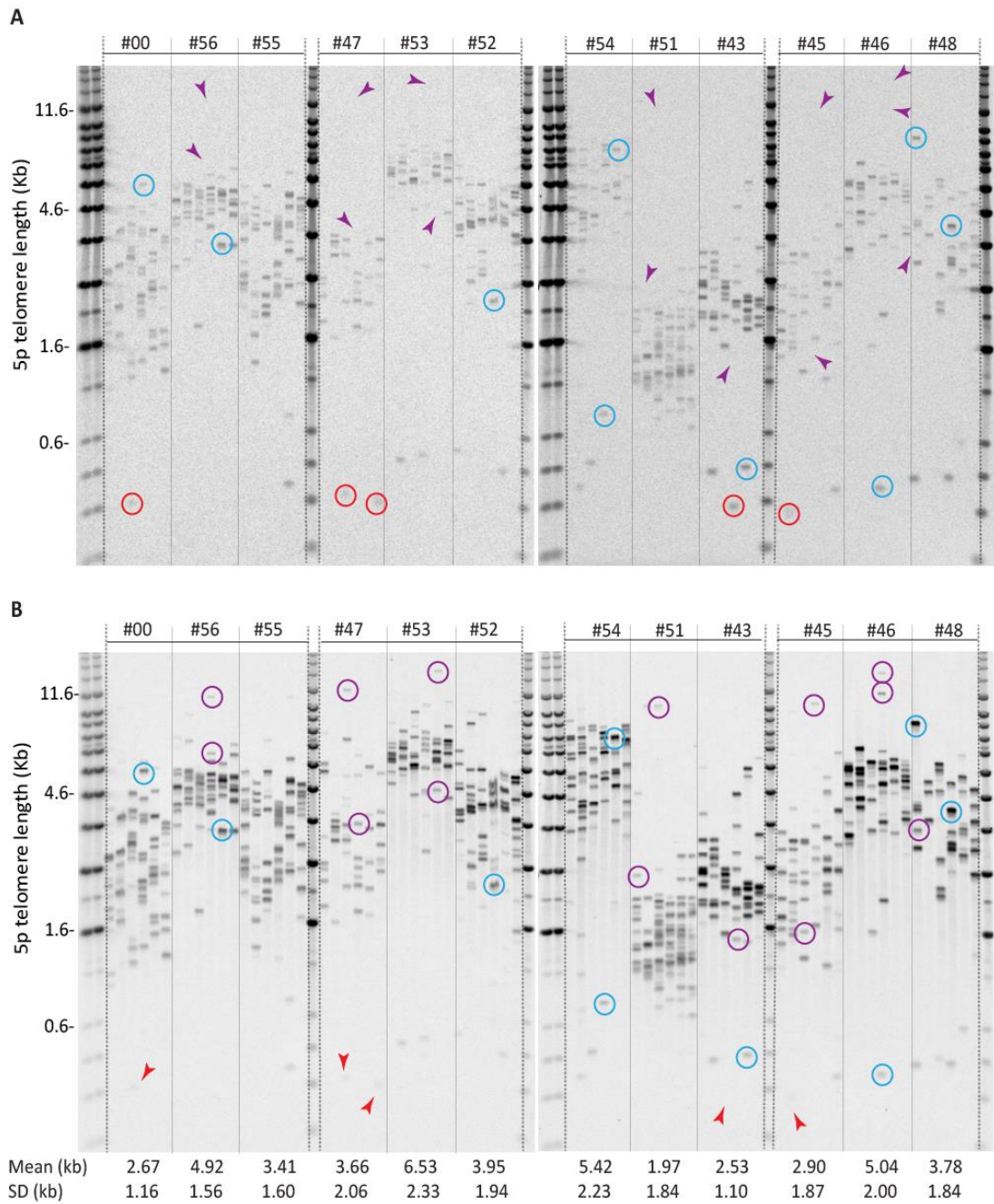


Figure 3.3. 5p STELA on 12 CLL patient samples.

STELA products were detected respectively with **(A)** the 5pA telomere adjacent probe and **(B)** the telomere probe. Some examples of the three different types of telomeres hybridisation pattern are shown with the circles (presence) and arrows (absence) of different colours. Type I: most of the telomeres, detected with both probes (blue). Type II: Very short telomeres (TL<1Kb) detected with the telomere adjacent probe but not the telomere probe (red). Type III: telomeres (TL>1Kb) detected with the telomere probe but not the 5pA probe (purple). Telomere mean and standard deviation (SD) are shown underneath. STELA products were generated using 1250pg DNA per reaction with the 5p5 primer.

3.4.2 Comparison of the telomere lengths at 5p, 17p and XpYp in CLL patients

To examine how the 5p TL profiles compared to other unique human telomeres, the TL in 57 CLL patients was measured using STELA at the 5p, 17p and XpYp telomeres.

The STELA reactions from 5p, 17p and XpYp from the same template DNA stock were analysed on the same gels (**Figure 3.4**). The TL for each individual telomere was measured and the distance from each primer to the start of the telomere was subtracted to obtain the real TL (5p=0.367Kb, 17p=0.311Kb and XpYp=0.408Kb). The mean TL, standard deviation (SD) and standard error (SE) at each chromosome end was calculated for 57 CLL patients (**Table 3.1**, **Figure 3.4** and **Supplementary Table 3**).

Table 3.1 Descriptive analysis at the 5p, 17p and XpYp telomeres for 57 CLL patients.

Telomere	5p	17p	XpYp
Mean TL (Kb)	3.180	2.917	2.511
SD (Kb)	1.164	1.075	0.977
SE (Kb)	0.154	0.142	0.129

Regression analysis was used to predict the relation of the TL among the distinct telomeres. Results from the comparison of the mean TL revealed a significantly positive correlation ($p < 0.0001$) as follows 5p-17p: $r^2 = 0.6453$, 5p-XpYp: $r^2 = 0.7220$ and 17p-XpYp: $r^2 = 0.7366$. These results indicated a positive association of the TL measured for the individual telomeres for each same patient (**Figure 3.5A-C**).

Telomere length analysis of 57 CLL patients samples revealed that the 5p telomere was significantly longer when compared to XpYp ($p < 0.05$). Similarly, the 17p telomere was significantly longer than XpYp ($p < 0.05$). The 5p and 17p mean TL were greater than XpYp by 0.699Kb and 0.406Kb, respectively. However, there were no significant differences between 5p and 17p (**Figure 3.5D**).

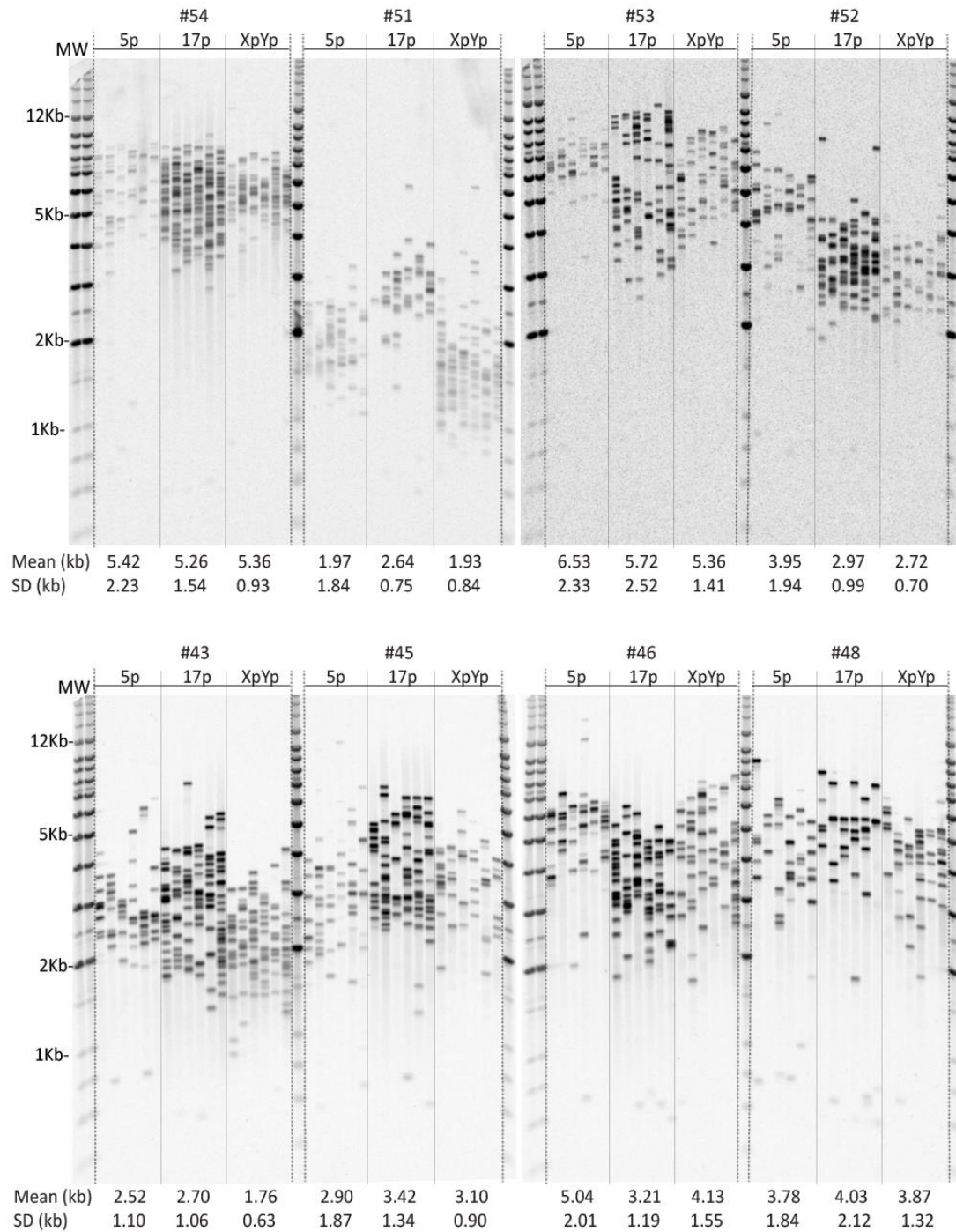


Figure 3.4. Comparison of the telomere lengths at 5p, 17p and XpYp.

Southern blot of four gels showing STELA at telomeres 5p 17p and XpYp for 8 of the 57 CLL samples. The mean and standard deviation are shown underneath. STELA products were generated using the following telomere-specific primers: 5p5, 17pseqrev1, XpYpE2, and detected with the telomere probe.

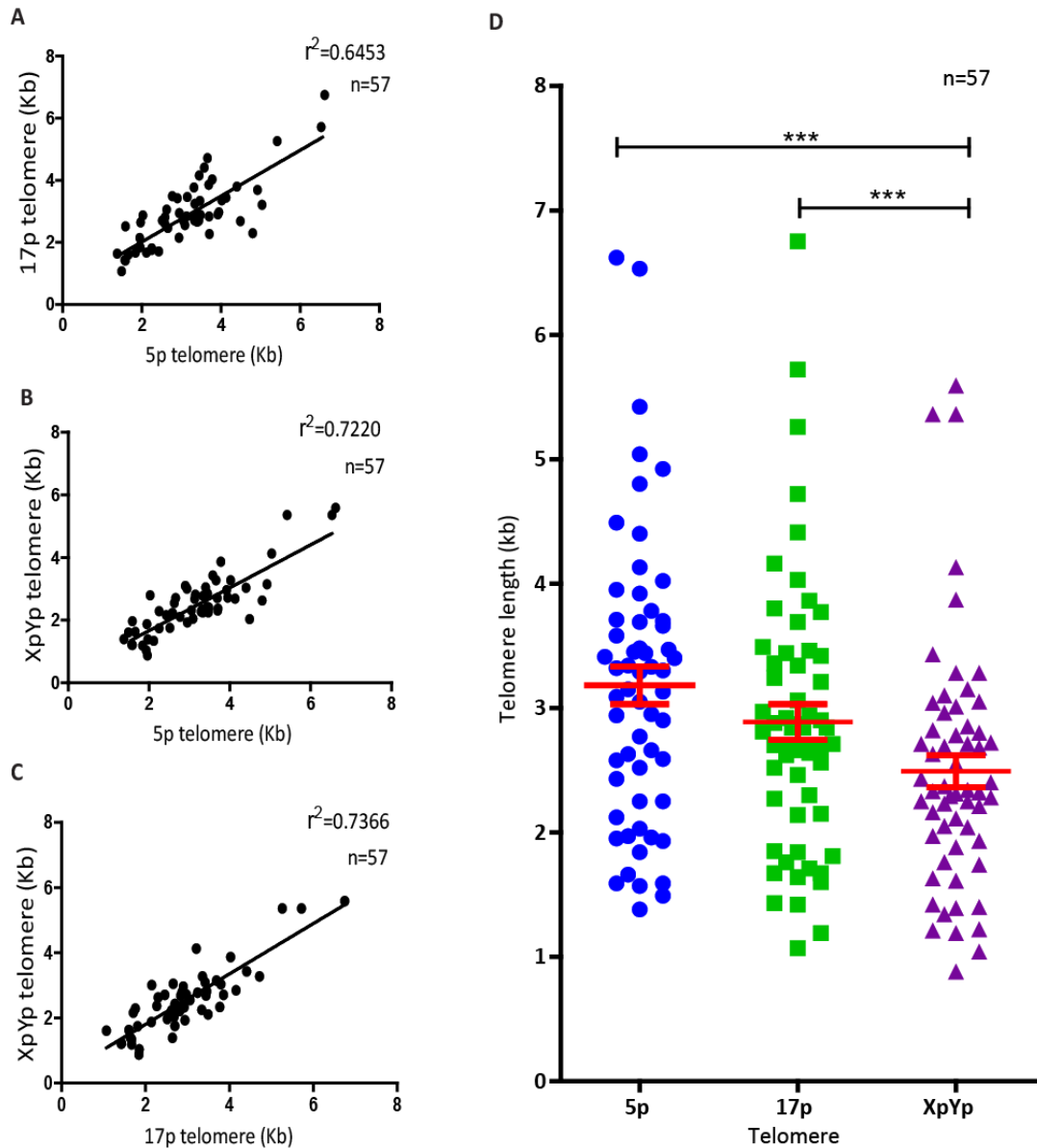


Figure 3.5. Comparison of the telomere length at 5p, 17p and XpYp for 57 CLL patients.

Correlation of the TL between the different chromosome ends **(A)** 5p-17p telomeres ($r^2=0.6453$), **(B)** 5p-XpYp telomeres ($r^2=0.7220$) and **(C)** 17p-XpYp telomeres ($r^2=0.7366$) with $p<0.0001$. **(D)** Scatter plot of mean TL at the 5p, 17p and XpYp chromosome ends. Samples did not follow Gaussian distribution (Shapiro-Wilk normality tests). Statistical analysis was performed with Friedman test $p<0.0001$ (non-parametric alternative to repeated measures ANOVA test) and Dunn's Multiple Comparison post-hoc test $p<0.05$. Standard error bars indicated in red.

3.4.3 The 5p telomere contains more variant repeats than the XpYp telomere

Telomere variant repeats (TVRs) such as TTGGGG, TGAGGG and TCAGGG (the most common ones) are located in the proximal regions of human telomeres interspersed with the canonical TTAGGG and can be found in variable amounts (Allshire *et al.*, 1989, Baird, 1995).

TVRs are non-functional telomeric repeats and may confound the real TL based on canonical TTAGGG repeats. To examine the nature of these differences on the 5p TL compared with XpYp, the variable amounts of the variant repeats TGAGGG and TCAGGG contained in the proximal regions of the 5p and XpYp telomeres were measured for 10 CLL patient samples using TVR-PCR (Baird, 1995). The mean TL was compared before and after correcting for the amount of TVRs. The most distal TVR position was subtracted from the mean TL (before subtracting the distance to the primer).

According to previous results, the 5p mean telomere was significantly longer (1.06Kb) than XpYp ($p=0.0019$) but significant differences disappeared (0.33Kb) after correcting the TL for the amount of TVR ($p=0.2685$) (**Figure 3.6** and **Supplementary Table 4**).

The 17p telomere was excluded from this study, as the presence of a constant non-specific PCR-artefact band prevented the measurement of the TVR. In addition, there were no differences between 5p and 17p TL, therefore this study focused on comparing the 5p to the XpYp.

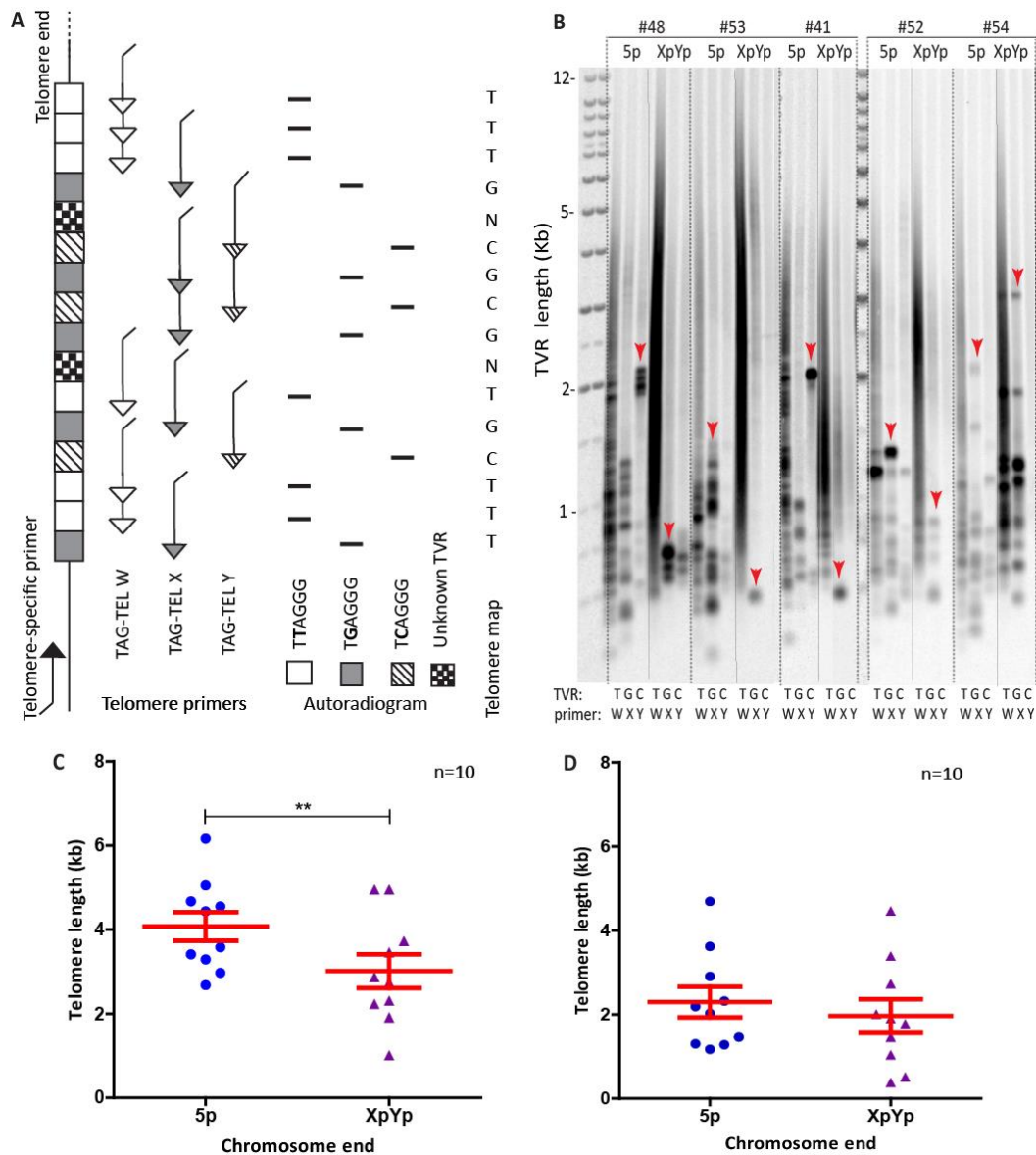


Figure 3.6. Comparison of the Telomere Variant Repeat at 5p and XpYp.

(A) The generation of a telomere map is achieved with the combination of the telomere-specific primers 5p5 and XpYpE2, together with the TAG-TelW, TAG-TelX, TAG-TelY, that amplify the telomere TTAGGG (white square), TVR TGAGGG (grey square) and TVR TCAGGG (hatched square), respectively. The products are resolved in a gel and the Southern blot is hybridised with a telomere probe. The telomere map can then be constructed. Unknown TVR are represented with a chequered square. Adaptation from Baird, 1995. **(B)** Examples of telomere maps by TVR-PCR from five of the 10 CLL samples analysed. The distribution of T (TTAGGG), G (TGAGGG), C (TCAGGG) and unknown can be read from the Southern blot of the gel at telomeres 5p and XpYp. Red arrows indicate the most distal position of the TVR detected for each individual at 5p compared to XpYp. **(C)** Comparison of the TL before and **(D)** after correcting by TVR. Shapiro-Wilk statistical normality tests ($p > 0.05$) indicated a Gaussian distribution. Paired Student's T-test for $n=10$ and $p < 0.05$, the p -values before and after correcting are 0.0019 and 0.2685, respectively. Standard error bars indicated in red.

3.4.4 5p telomere fusions are detected with the 5p8:17p6:XpYpM:16p1:21q1 telomere fusion assay in human cell lines and CLL patients

3.4.4.1 Development of the telomere fusion assay at the 5p chromosome end

To examine whether the 5p telomere was dysfunctional and fusogenic, the telomere fusion assay was adapted to include the 5p telomere. Selection of primers within 6kb of the 5p telomere was based on restricted specificity of 5p subtelomeric sequence (**Figure 3.7**, **Figure 3.8** and **Supplementary Table 5**).

A CpG island was observed within the 5p subtelomeric DNA as shown in **Figure 3.9** using the ENDEMO GC content calculator tool (<http://www.endmemo.com/bio/gc.php>). The CpG island was a region of ~922bp with 78.31% GC content. It can be difficult to amplify GC-rich sequence because of the formation of secondary structures that resist denaturation and impede primer annealing. Therefore primer combinations were located at either side avoiding and including the CpG island.

Each of the 5p pair of primers was tested in a conventional gradient PCR reaction (annealing temperature 58-68°C) for the HEK293 cell line, a CLL patient sample and a negative control (ddH₂O). Only the primer combinations A, E, F and G that excluded the CpG island at 5p were able to generate amplicons. (**Figure 3.8D**).

The optimal T_m, indicated with the amplification of a PCR product of expected size and the lack of non-specific PCR amplicons, was selected for each primer combination (**Figure 3.8D** and **Supplementary Table 6**).

Following amplification and purification of the PCR products from the agarose gel, DNA was Sanger sequenced and aligned to the human genome to verify the specificity of the primers. The resulting sequence was specific to the 5p subtelomere.

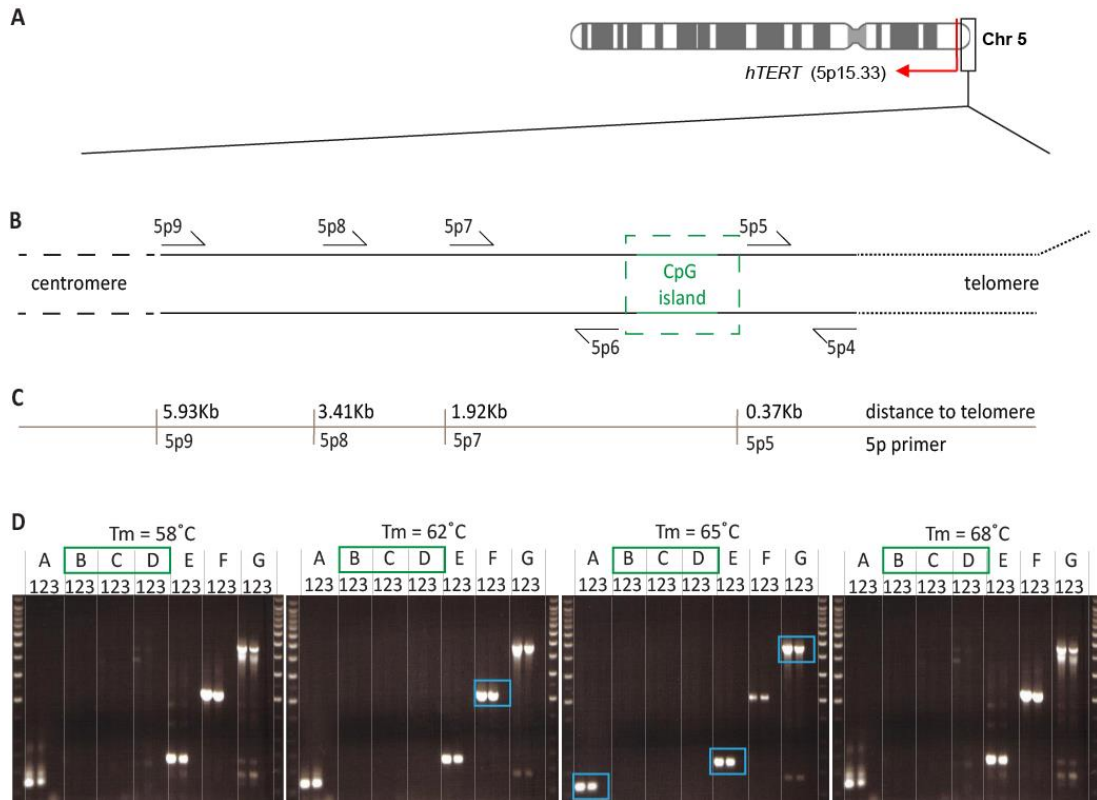


Figure 3.8. Primer design at the 5p chromosome end.

(A) Chromosome 5. The 5p telomere is highlighted by a black box and the location of the *hTERT* gene locus is indicated by a red arrow. (B) A representation of the 5p subtelomeric region with the location of primers and (C) distance to the start of the telomere. GC-rich DNA sequence is highlighted by a green box and highlights primer combinations at each side. (D) Agarose gels following gradient PCR (Annealing temperature 58-68°C) with different primer combinations (A: 5p5+5p4 (0.34Kb), B: 5p7+5p4 (1.89Kb), C: 5p8+5p4 (3.37Kb), D: 5p9+5p4 (5.90Kb) E: 5p6+5p7 (0.64Kb), F: 5p8+5p6 (2.12Kb) and G: 5p9+5p6 (4.64Kb)). A was used as the 5p telomere-adjacent DNA probe for 5p STELA. F and G as 5p fusion probes in the fusion assay. The most appropriate T_m for each primer combination is highlighted by a blue box. 1: HEK293 cell line, 2: CLL patient sample and 3: negative control (ddH₂O). Green box indicates primer combinations that span the CpG island.

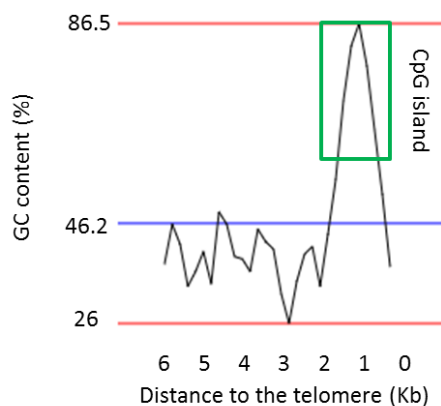


Figure 3.9. GC distribution at the 5p subtelomere.

ENDMEMO GC was used for creation of data on GC content in Human Feb. 2009 (GRCh37/hg19) 5p subtelomeric DNA sequence (0-6Kb). The GC island (922bp) had a 78.31% average GC content. The blue line indicates the average GC content (%). Telomere sequence (TTAGGG) was not included in the figure (GC content ~50%).

3.4.4.2 Optimisation: testing different telomere fusion primer combinations

The single-molecule telomere fusion assay originally developed in our laboratory, consists of four primers, two of which target unique telomeres (17p and XpYp) and two that target several distinct telomeres (16p family: 16p, 1p, 9p, 12p, 15q, XqYq and 2q14 interstitial telomeric locus, and 21q family: 21q, 1q, 2q, 5q, 6q, 6p, 8p, 10q, 13q, 17q, 19p, 19q, 22q and the 2q13 interstitial telomeric locus), amplifying a total of 23 telomeres (Capper *et al.*, 2007, Letsolo *et al.*, 2010).

To investigate whether the 5p telomere is fusogenic, which could result in genomic instability in this chromosome arm where *hTERT* is located, the telomere fusion assay was adapted to include the 5p telomere. To maximise the number of detectable fusions a systematic analysis of fusion amplicons generated using different 5p telomere-specific primers in combination with our group's fusion assay was conducted.

Each of the 5p primers was tested independently in the fusion PCR assay, on their own and in combination with 17p6, XpYpM, 16p1 and 21q1 primers with DNA extracted from the human HEK293 cell line and a CLL patient sample. The hybridisation pattern with the different probes revealed the specificity of the 5p primers. With the 5p primer alone, only products with the 5p probe and not the other probes were detected. In contrast, with the 17p6:XpYpM:16p1:21q1, PCR amplicons were hybridised with the other probes, but not with the 5p probe (**Figure 3.10** and **Figure 3.11**).

Southern hybridisation with the 5p9G probe revealed a substantial amount of PCR artefacts, non-fusion amplicons resulting from the inclusion of the 5p9 primer (**Figure 3.10**). By performing *in-silico* PCR with the different primers involved (5p9, 17p6, XpYp, 16p1 and 21q1), the non-specific PCR background signal was resolved to be an off-target event created by the 5p9 primer in combination with the 21q1 primer (**Table 3.2**). Compared to 5p9, 5p8 resulted in efficient amplification of fusion molecules with only a single characteristic constitutive background band (non-fusion amplicons) of 1.4kb (**Figure 3.11** and **Figure 3.12A**). Therefore, the 5p8 primer that was located 3.4Kb away from the start of the telomere, was selected and included in the fusion assay.

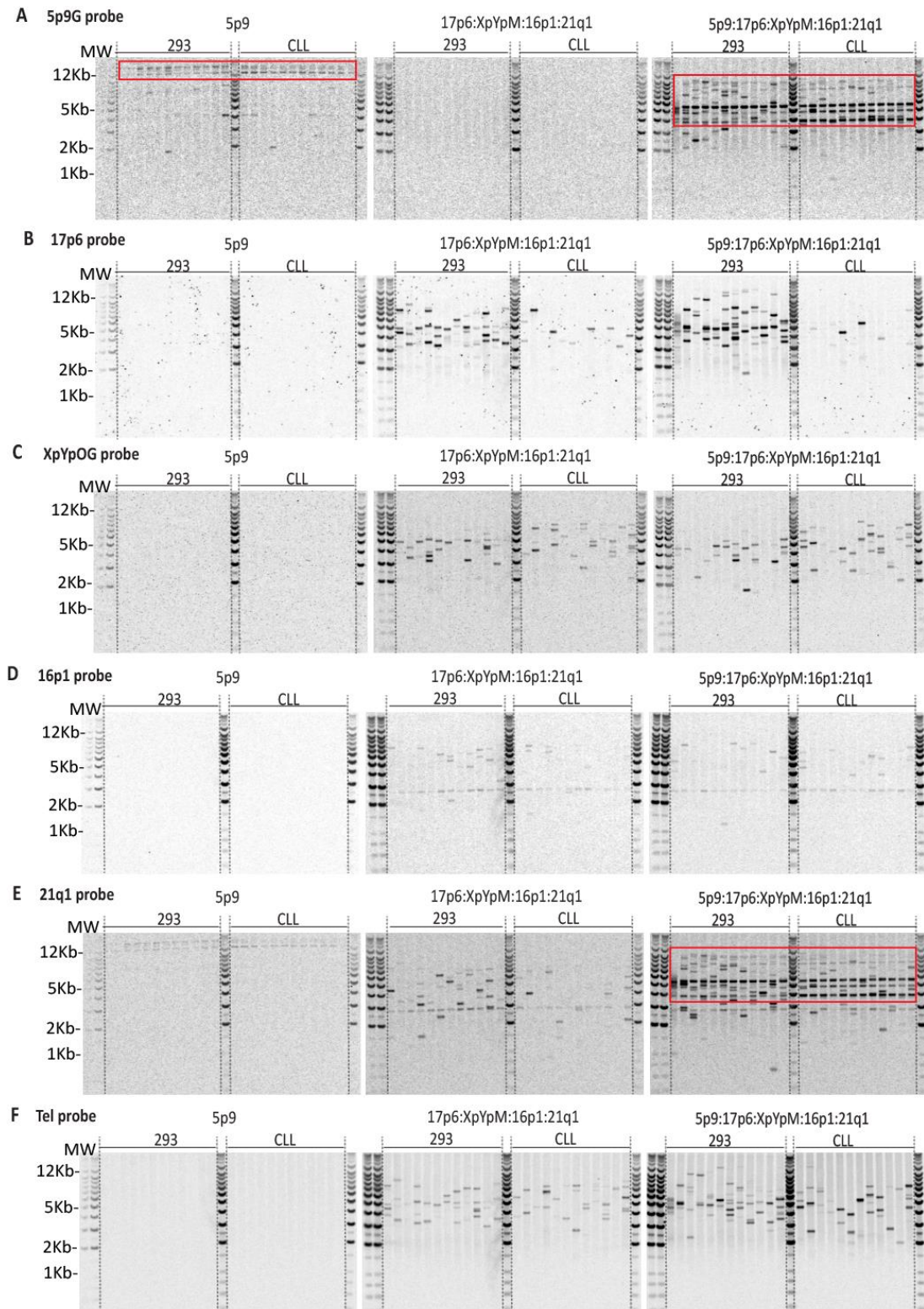


Figure 3.10. Testing the 5p9 primer into the fusion assay.

Telomere fusion assay with 5p9, 17p6:XpYpM:16p1:21q1 and 5p9:17p6:XpYpM:16p1:21q1 using the HEK293 cell line and a CLL patient sample at T_m 62°C. Southern blots were hybridised with (A) 5p9G, (B) 17p6, (C) XpYpOG, (D) 16p1, (E) 21q1 and (F) Telomere probes. PCR non-specific products produced with the 5p9 and 21q1 primers highlighted with a red box.

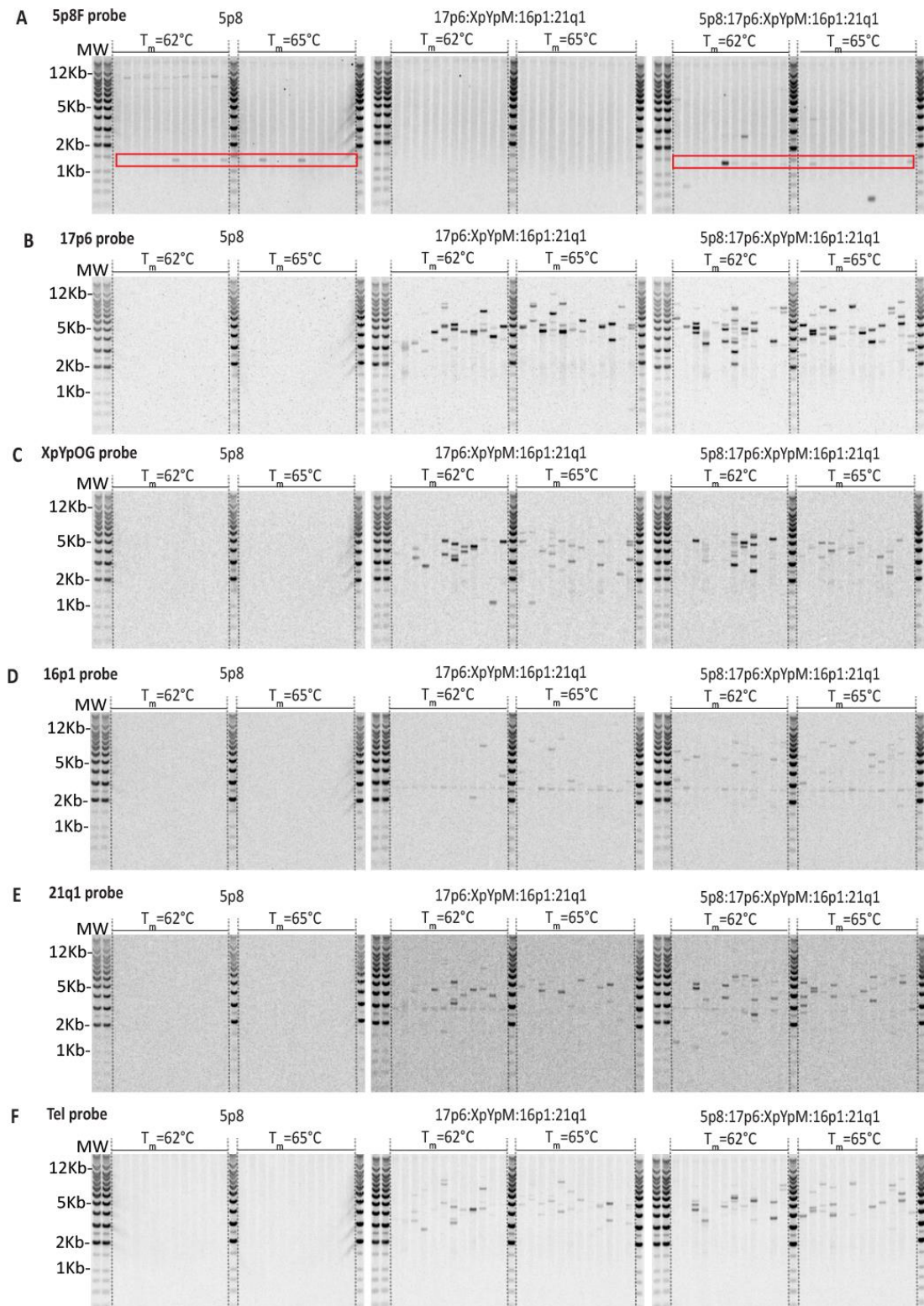


Figure 3.11. Testing the 5p8 primer into the fusion assay.

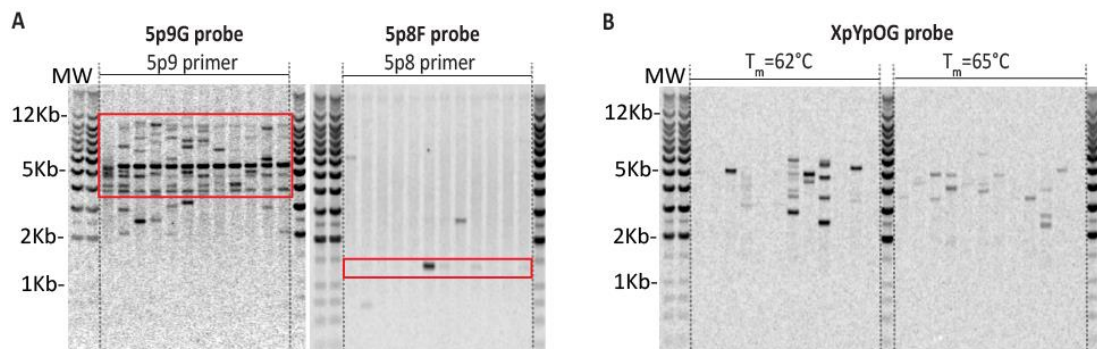
Telomere fusion assay with 5p8, 17p6:XpYpM:16p1:21q1 and 5p8:17p6:XpYpM:16p1:21q1 using the HEK293 cell line at T_m 62°C and 65°C. Southern blots were hybridised with (A) the 5p8F, (B) 17p6, (C) XpYpOG, (D) 16p1, (E) 21q1 and (F) Telomere radiolabelled probes. PCR non-specific products produced with the 5p8 primer highlighted with a red box.

Table 3.2. *In-silico* PCR results using the 5p9 and 21q1 primers.

Primers	Location	Product size (Kb)
5p9 + 21q1	chr1:249226143+249236498	10.356
	chr2:243155772-243164493	8.722
	chr5:180895308+180904773	9.466
	chr8:158924-167369	8.446
	chr10:135510156+135521468	11.313
	chr19:249134-255942	6.809
	chr19:59104902+59115960	11.059
	chr21:48105249+48116107	10.859

The PCR annealing temperature was increased from 62°C to 65°C to reduce non-specific products resulting from unspecific binding of the primer. However, the 1.4Kb band could not be eliminated and the number and intensity of the telomere fusion amplicons decreased at 65°C (**Figure 3.11** and **Figure 3.12B**).

Altogether, a T_m of 62°C annealing temperature using the 5p8 primer was selected for subsequent experiments.

**Figure 3.12. Highlights from the optimisation of the 5p8:17p6:XpYpM:16p1:21q1 fusion assay.**

Comparison of amplicons detected after performing the telomere fusion assay (**A**) with the 5p9 and the 5p8 primer. Southern blots were hybridised with the 5p9G and 5p8F, respectively. (**B**) Comparison of results obtained using the 5p8 primer at T_m 62°C and 65°C. Southern blot was hybridised with the XpYp probe. PCR non-specific products highlighted with a red box.

3.4.4.3 The 5p8 PCR artefact is generated by non-specific binding of the 5p8 primer

Characterisation of the 5p8 constitutive background band (1.4Kb) was completed by reamplification of the band using the 5p8 primer. The DNA was subsequently purified from the agarose gel and Sanger sequenced. As shown in **Figure 3.13**, the 1.4Kb PCR product was generated by 7bp at the 3' end of the 24bp-5p8 primer bound 1.377Kb upstream from the 24bp-binding site.

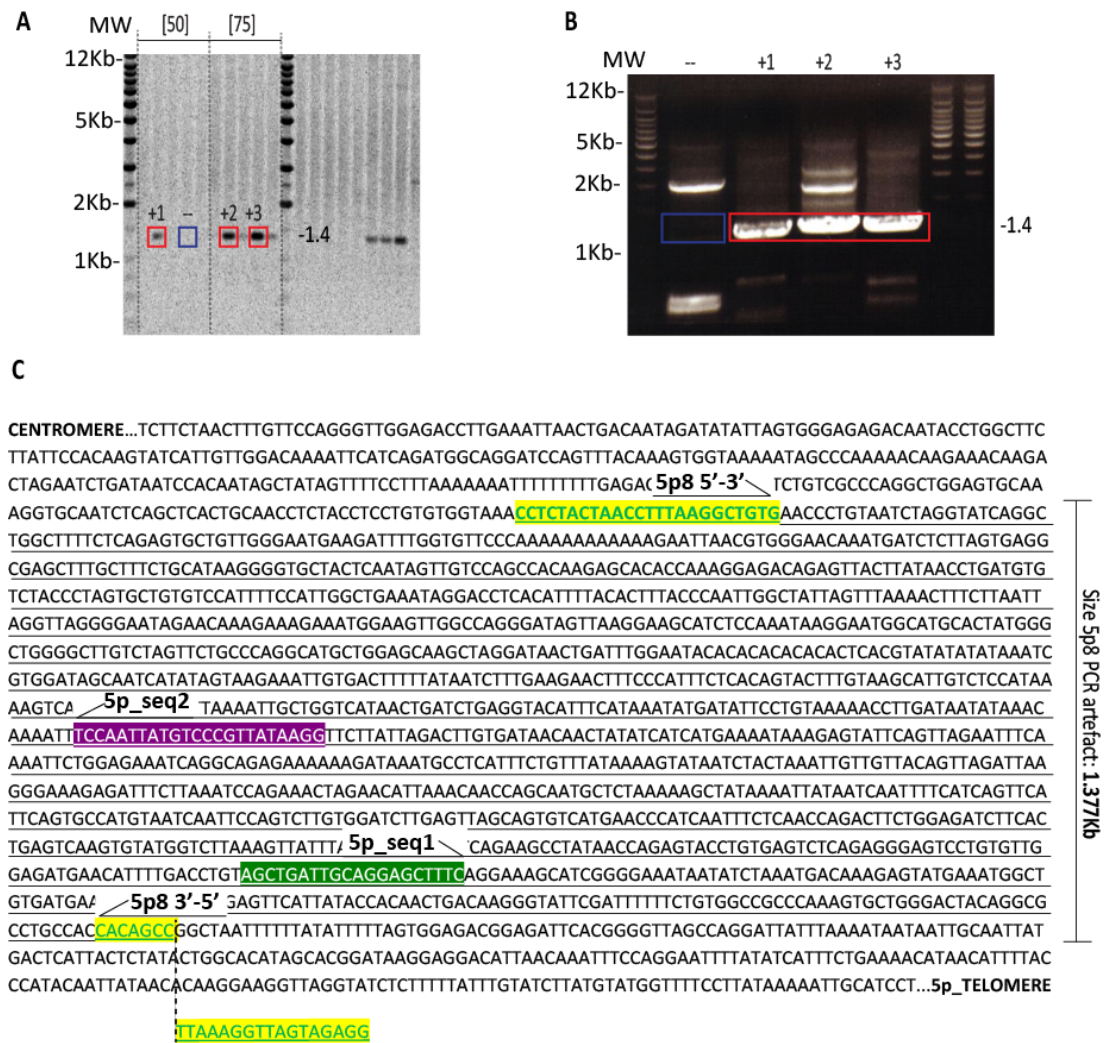


Figure 3.13. Characterisation of the 5p8 PCR artefact.

(A) The Southern blot was hybridised with the 5p8F probe following the 5p8:17p6:XpYpM:16p1:21q1 telomere fusion assay on a CLL patient sample. (B) An agarose gel depicting the reamplification of the PCR artefact using only the 5p8 primer. The 5p8 PCR artefact is highlighted by a red box and the negative control is indicated by a blue box. (C) Nucleotide sequence product and primers used to reveal the cause of the artefact. Size 1.377Kb (~1.4Kb).

3.4.4.4 The inclusion of the 5p telomere in the fusion assay increases the detection of telomere fusion amplicons

The fusion amplification efficiency was compared using the new 5p8:17p6:XpYpM:16p1:21q1 combination with the former 17p6:XpYpM:16p1:21q1 method (without 5p8) using DNA derived from HEK293 cell line. The frequency of fusions (relative number of fusions per diploid genome) was calculated after detection of fusion events with each different radiolabelled probe.

It is important to note that the frequency of fusions is a useful value that will be used throughout this project but only as a relative representation as this underestimates the frequency of fusions genome-wide.

$$\text{Frequency of fusions} = \frac{\# \text{ Fusion event}}{\# \text{ Diploid genome}}$$

$$\# \text{ Diploid genome} = \frac{1 \text{ diploid genome}}{6 \text{ pg DNA}} \times \frac{10^3 \text{ pg DNA}}{1 \text{ ng DNA}} \times \frac{100 \text{ ng DNA}}{1 \text{ reaction}} \times \# \text{ reactions}$$

Example: Three 5p fusion events detected in 12 PCR reactions containing 100ng each of gDNA derived from HEK293 cell line is a frequency of 1.5×10^{-5} fusion events per diploid genome.

$$\text{Frequency of 5p fusions} = \frac{\# \text{ Fusion event}}{\# \text{ Diploid genome}} = \frac{3}{2 \times 10^5} = 1.5 \times 10^{-5}$$

$$\# \text{ Diploid genome} = \frac{1 \text{ diploid genome}}{6 \text{ pg DNA}} \times \frac{10^3 \text{ pg DNA}}{1 \text{ ng DNA}} \times \frac{100 \text{ ng DNA}}{1 \text{ reaction}} \times 12 \text{ reactions} = 2 \times 10^5$$

Using the 5p8:17p6:XpYpM:16p1:21q1 fusion assay compared with the 17p6:XpYpM:16p1:21q1 (excluded 5p8) resulted in an increase of fusions frequency from 3.9×10^{-4} to 5.7×10^{-4} (**Figure 3.14** and **Table 3.3**). Therefore, the 5p8:17p6:XpYpM:16p1:21q1 fusion PCR assay was selected to detect the maximum number of fusion events present for each sample.

The consecutive detection of the fusion products with different radiolabelled probes identified a 5p-17p inter-chromosomal fusion event and 2 possible 5p sister chromatid fusions in HEK293 cell line. Neither of them contained substantial telomere repeats as they were not hybridised by the telomere probe (**Figure 3.14A**).

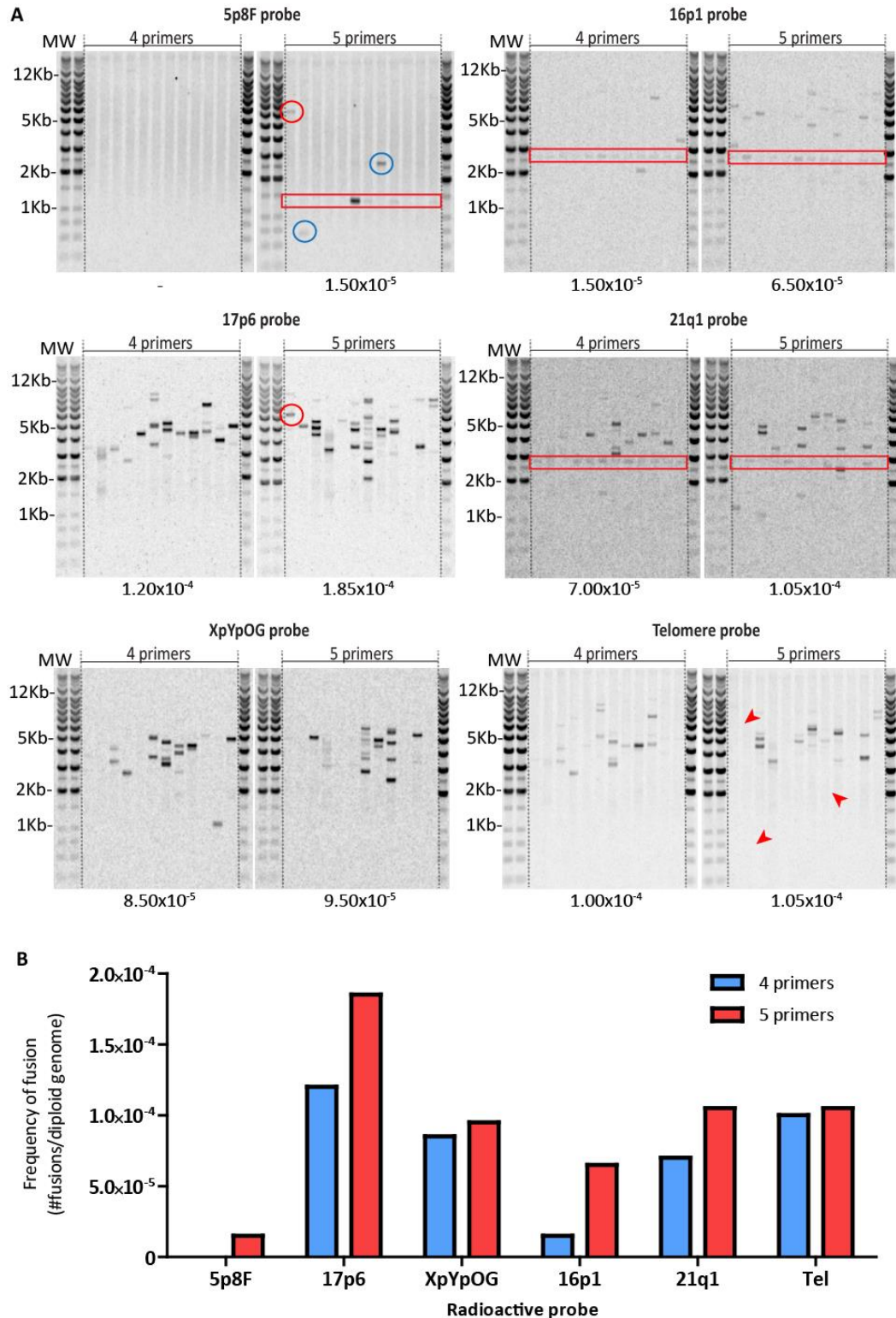


Figure 3.14. Comparison of the distinct telomere fusion assays for HEK293 cell line at 62°C.

(A) Fusion products generated by the 17p6:XpYpM:16p1:21q1 (4 primers) and 5p8:17p6:XpYpM:16p1:21q1 (5 primers) telomere fusion assays were detected consecutively with the 5p8F, 17p6, XpYpOG, 16p1, 21q1 and telomere probe. Estimated frequency of fusion is listed underneath each image. Highlighted in the red circle an example of a 5p-17p inter-chromosomal fusion event. Two possible 5p-5p sister chromatid fusions, since they only hybridise with the 5p

probe, highlighted in a blue circle. The absence of the fusion events containing 5p when hybridising with the telomere probe are highlighted by red arrows. **(B)** Comparison of the frequency of fusion detected between the 17p6:XpYpM:16p1:21q1 (4 primers) and the 5p8:17p6:XpYpM:16p1:21q1 (5 primers) fusion assays.

Table 3.3. Frequency of fusions detected with the distinct telomere fusion assays.

Telomere fusion assay	5p8F	17p6	XpYpOG	16p1	21q1	Telomere
17p6:XpYpM:16p1:21q1	0.00x10 ⁺⁰	1.20x10 ⁻⁴	8.50x10 ⁻⁵	1.50x10 ⁻⁵	7.00x10 ⁻⁵	1.00x10 ⁻⁴
5p8:17p6:XpYpM:16p1:21q1	1.50x10 ⁻⁵	1.85x10 ⁻⁴	9.5x10 ⁻⁵	6.50x10 ⁻⁵	1.05x10 ⁻⁴	1.05x10 ⁻⁴

3.4.5 Characterisation of the 5p telomere in a HCT116^{DN-hTERT} clone progressing through a telomere-driven crisis.

To study telomere erosion and progression through a telomere-driven crisis *in vitro*, our group transfected wild type (WT) HCT116 cell line with a DN-*hTERT* construct to inhibit telomerase expression. Telomere fusions were detected during crisis, and escape was associated with amplification of the 5p chromosome arm including the *hTERT* locus and upregulation of telomerase (Jones *et al.*, 2014). Two examples for which this was observed included HCT116^{DN-hTERT} WT clone 11 and LIG3^{-/-mL3+nuc. LIG3} clone 15 (**Figure 3.15**) (Jones *et al.*, 2014).

To investigate whether telomere erosion, dysfunction and fusion at the 5p telomere was associated with the amplification of 5p and reactivation of telomerase, 5p STELA and the 5p8:17p6:XpYpM:16p1:21q1 fusion assay were used in HCT116^{DN-hTERT} WT clone 11 and LIG3^{-/-mL3+nuc. LIG3} clone 15.

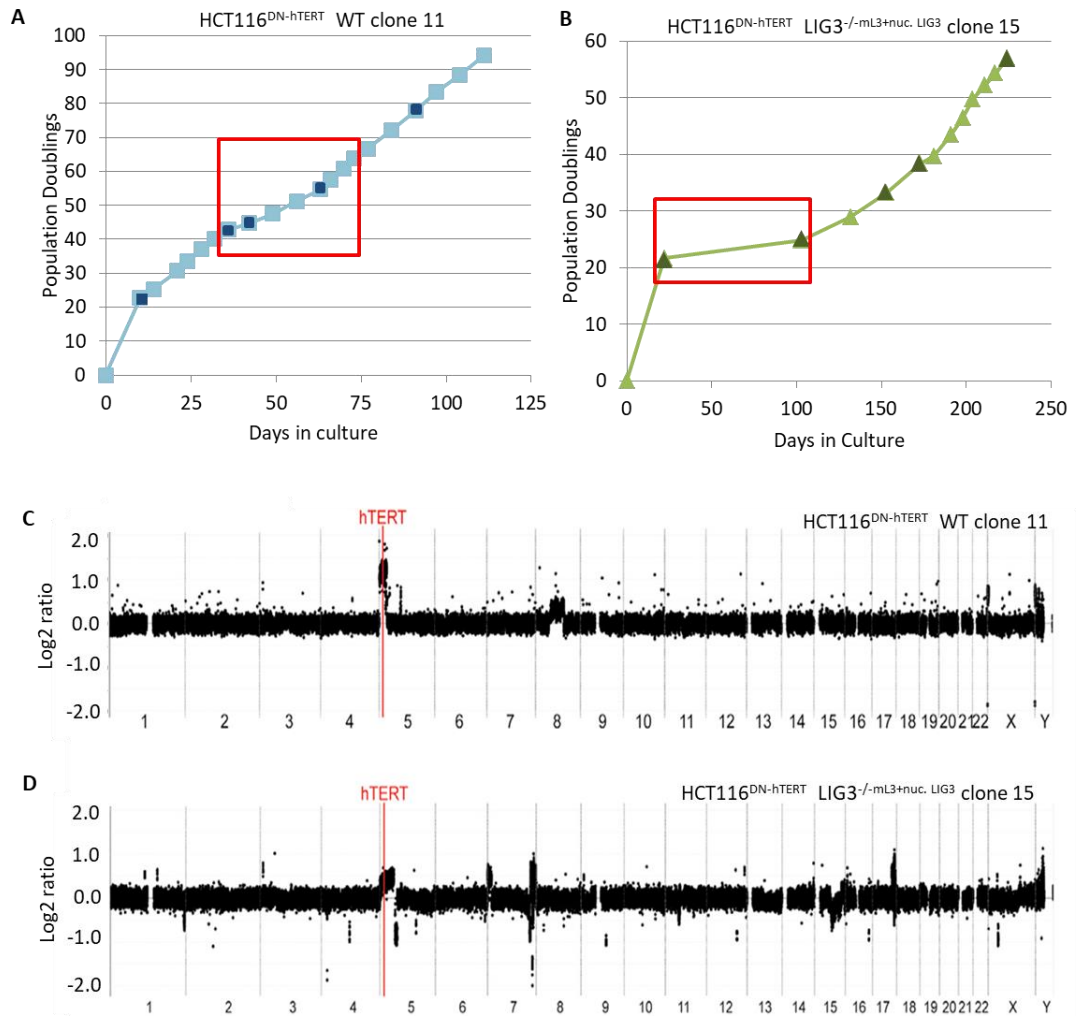


Figure 3.15. HCT116^{DN-hTERT} cell lines.

Growth curve for HCT116^{DN-hTERT} (A) WT clone 11 and (B) LIG3^{-/-mL3+nuc. LIG3} clone 15. Red box highlights telomere-driven crisis. Darker time points indicate samples selected for STELA and fusion analysis. (C) Array CGH on samples taken before and after crisis showing amplification of the 5p chromosome end where *hTERT* is located after escape of telomere-driven crisis for HCT116^{DN-hTERT} WT clone 11 and (D) LIG3^{-/-mL3+nuc. LIG3} clone 15. Adapted from Jones *et al.*, 2014.

3.4.5.1 Characterisation of the 5p telomere dynamics using STELA

STELA at the 5p and XpYp telomeres revealed the gradual shortening of the TL as cells approached telomere-driven crisis. This was followed by a gradual elongation of the TL after escape of crisis that was especially observed using $LIG3^{-/-mL3+nuc. LIG3}$ clone 15 (**Figure 3.16** and **Supplementary Table 7**).

For WT clone 11, the 5p telomere eroded 1.27Kb during 56 days and elongated 0.15Kb during 25 days (one time point available after crisis) (**Figure 3.16A**). Erosion could not be calculated for $LIG3^{-/-mL3+nuc. LIG3}$ clone 15 as the first time point available was already in crisis but the 5p telomere elongated 1.92Kb during 199 days after escape of crisis (**Figure 3.16B**). Similar trend was observed on the XpYp telomere for both populations (**Figure 3.16CD**).

Moreover, the TL profile appeared very homogeneous for WT clone 11. In contrast, a very heterogeneous TL distribution could be observed after crisis in $LIG3^{-/-mL3+nuc. LIG3}$ clone 15. Particularly after day 153 were two or more distinct TL clusters could be observed (**Figure 3.16**).

The comparison of the 5p to the XpYp mean TL at different time points for WT clone 11 (**Figure 3.17A**) and $LIG3^{-/-mL3+nuc. LIG3}$ clone 15 (**Figure 3.17B**) revealed that the 5p telomere was significantly longer than the XpYp telomere (Mann Whitney non-parametric T-test $p < 0.05$). The 5p telomere was 0.934Kb greater than the XpYp telomere (mean differences for the 10 time points). A significant direct correlation was also observed ($r^2 = 0.7047$, $p = 0.0024$, $n = 10$). These results were accordant with the comparison of the 5p with the XpYp TL in CLL patient samples before TVR correction (**sections 3.4.2** and **3.4.3**).

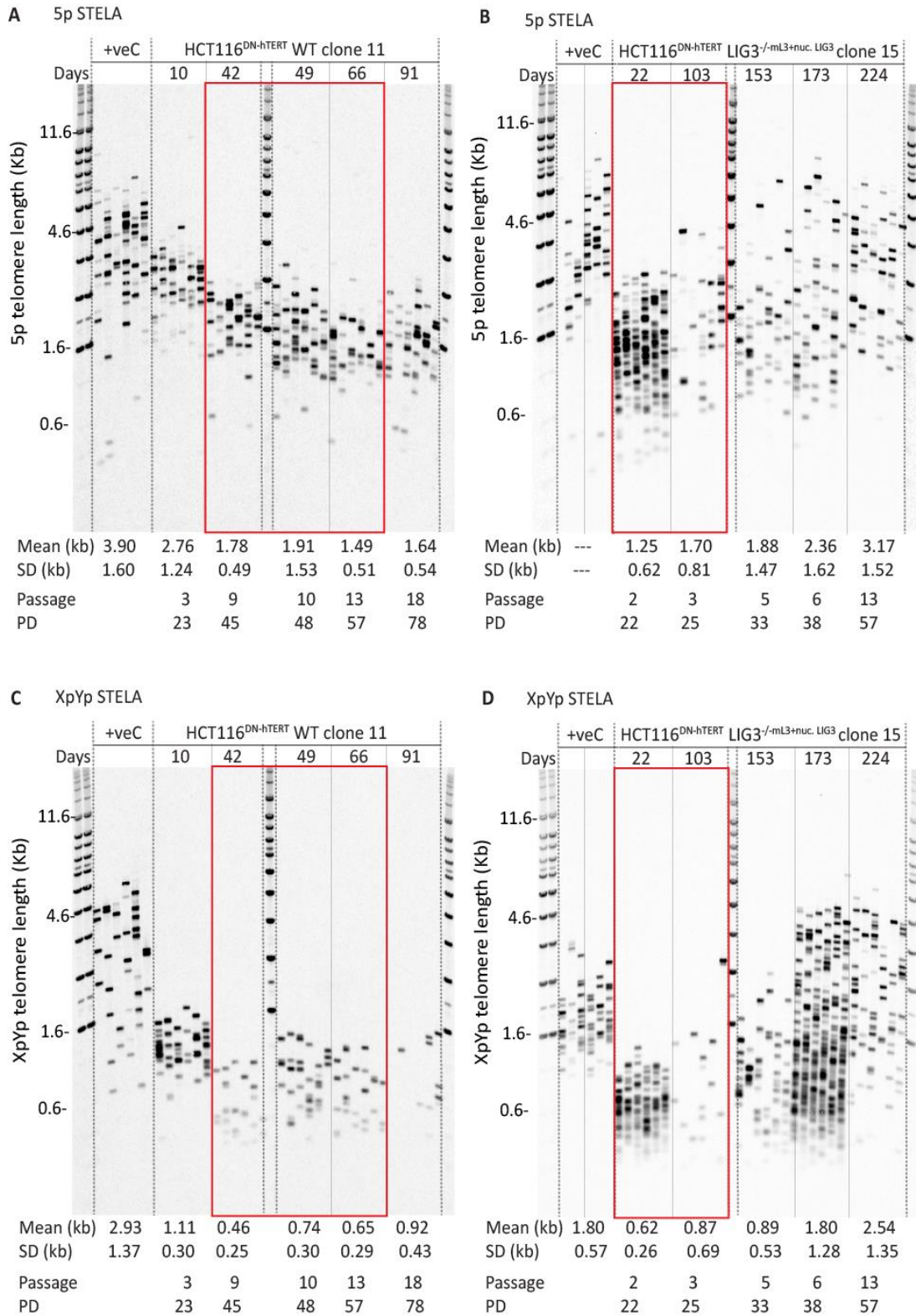


Figure 3.16. Characterisation of the 5p telomere using STELA.

(A, B) 5p and (C, D) XpYp STELA on (A, C) HCT116^{DN-hTERT} WT clone 11 and (B, D) LIG3^{-/-}mL3+nuc.LIG3 clone 15. Mean TL, SD, passage and population doubling (PD) are indicated underneath. Time points during crisis are highlighted by a red box. Southern blots were hybridised with the telomere probe.

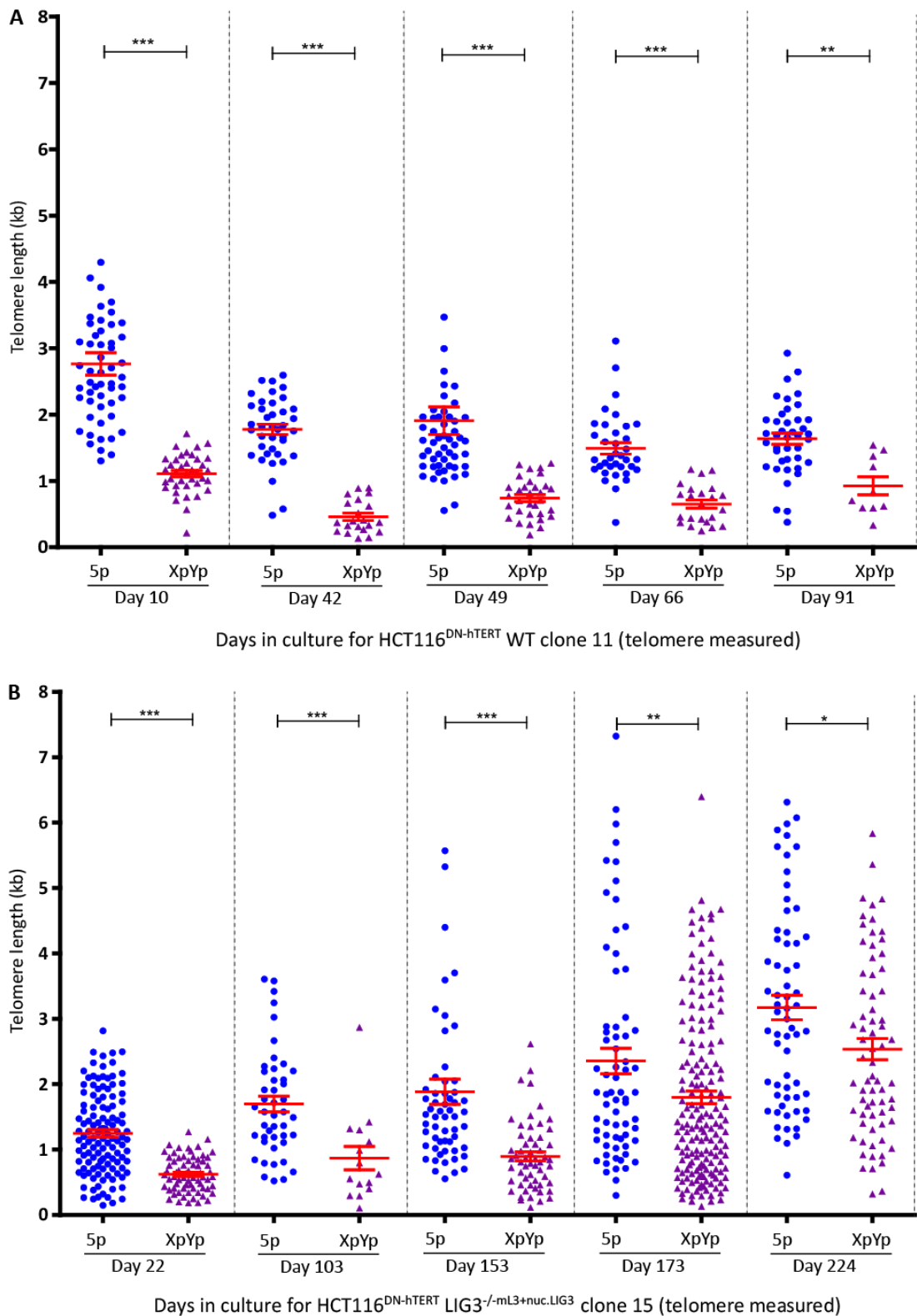


Figure 3.17. Comparison of the 5p to the XpYp telomere.

Graphic representation of STELA at the 5p compared to the XpYp telomere for **(A)** HCT116^{DN-hTERT} WT clone 11 and **(B)** LIG3^{-/-}-mL3+nuc.LIG3 clone 15. Statistical analysis was performed with Mann Whitney non-parametric T-test $p < 0.0001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*).

To assess whether an amplification or deletion of the 5p telomere occurred, the ratio of molecules amplified for 5p was compared to XpYp. **Figure 3.18** illustrates the 5p/XpYp ratio calculated for WT clone 11 (**A**) and $LIG3^{-/-mL3+nuc.LIG3}$ clone 15 (**B**) at each time point and subsequently normalised to the first time point (prior to crisis). WT clone 11 revealed a consistent normalised 5p/XpYp ratio for all days in culture except for day 91 (2.94) where a reduction of XpYp telomeres amplified was observed (**Figure 3.18A**). There were no significant differences for any of the groups except for XpYp day 10 compared to 91 (two-way ANOVA, $p < 0.05$). In contrast, for $LIG3^{-/-mL3+nuc.LIG3}$ clone 15, a reduction in the amount of 5p telomeres after crisis (day 153-224) was observed as indicated by normalised ratios 0.58, 0.22, 0.52 (**Figure 3.18B**). However, in general there were no significant differences (two-way ANOVA, $p > 0.05$). Altogether there were no clear differences in telomere dynamics in terms of length change or LOH at 5p compared to XpYp in these cells transiting crisis.

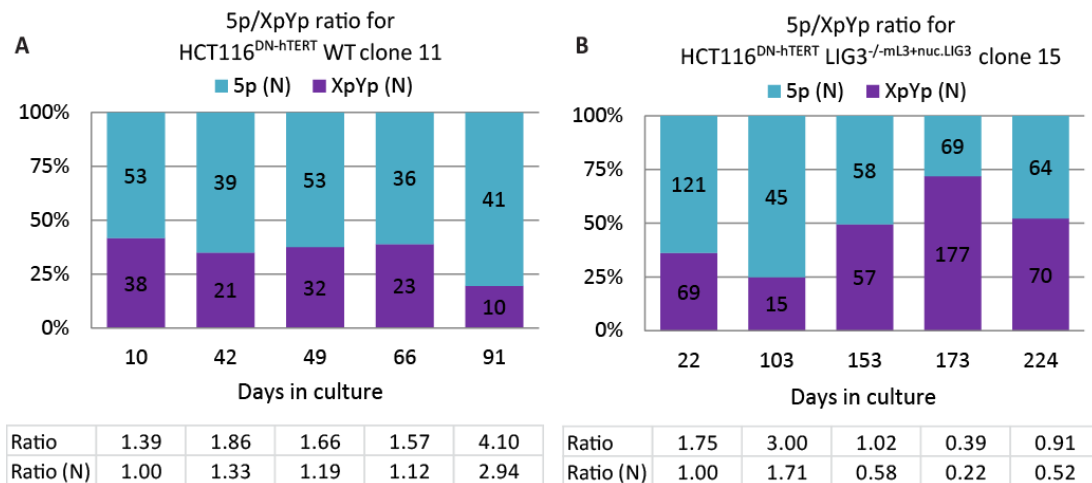


Figure 3.18. Comparison of the 5p to the XpYp molecules amplified using STELA.

5p/XpYp ratio and normalised to first time point ratio for (**A**) HCT116^{DN-hTERT} WT clone 11 and (**B**) $LIG3^{-/-mL3+nuc.LIG3}$ clone 15 indicated underneath.

Telomere instability at the 5p telomere of HCT116^{DN-hTERT} WT clone 11 and $LIG3^{-/-mL3+nuc.LIG3}$ clone 15 was measured at different time points before and after crisis using the optimised 5p8:17p6:XpYpM:16p1:21q1 fusion assay. Hybridisation with the distinct telomere specific probes revealed potential 5p-5p sister chromatid fusion events as well as possible 5p inter-chromosomal fusions for WT clone 11 and $LIG3^{-/-mL3+nuc.LIG3}$ clone 15. However, these events were rare compared with the other telomeres included in the fusion assay. Overall, a higher frequency of fusions was observed during crisis than before or after crisis (**Figure 3.19**).

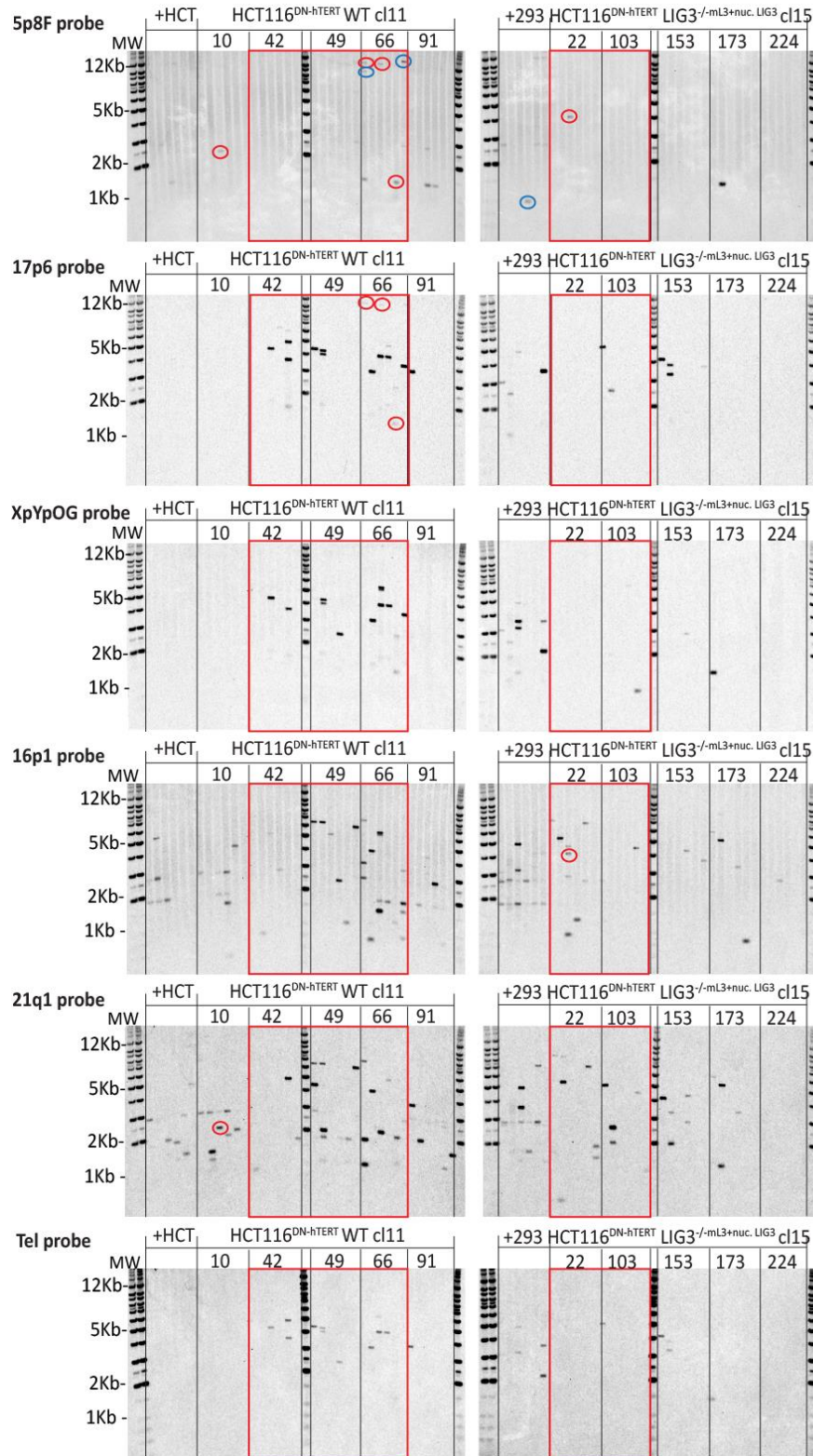


Figure 3.19. 5p8:17p6:XpYpM:16p1:21q1 fusions on HCT116 DN-hTERT cell lines.

Telomere fusion analysis with the 5p8:17p6:XpYpM:16p1:21q1 combination for HCT116 WT clone 11 and LIG3^{-/-}mL3+nuc. LIG3⁺ clone 15. Southern blots were hybridised with the 5p8F, 17p6, XpYpOG, 16p1, 21q1 and telomere probes. Days in culture are indicated at the top. Time points during crisis are indicated by a red box. 5p fusion events are highlighted by a blue circle.

3.4.6 Telomere dysfunction and fusion in CLL, including the 5p telomere

Genomic instability at the 5p chromosome arm, including *hTERT*, has been observed in some CLL patients and associated with increased telomerase expression (Nagel *et al.*, 2010, Schilling *et al.*, 2013, Salaverria *et al.*, 2015). Furthermore, fusion events can be identified in patient CLL-B cells with short telomeres (TL<3.81Kb) and that TL range can identify a subgroup of patients with a more aggressive disease (Lin *et al.*, 2010, Lin *et al.*, 2014).

To identify CLL patients with the highest frequency of telomere fusions and investigate rare telomere fusion events, patients with short telomeres were selected. This would also allow determining whether the 5p telomere is dysfunctional and fuses in CLL, which has the potential to drive the instability observed at 5p.

The first datasets collected were samples from CLL patients undergoing consultation at the University Hospital of Wales (UHW) of varying TL: 26 patient samples presented TL <3.81Kb and 7 presented TL between 3.81-5.59Kb.

Dr Kevin Norris analysed XpYp TL using a high-throughput version of STELA (HT-SELA) (Norris *et al.*, manuscript in preparation) on a cohort of 209 and 276 CLL patient samples from the LRF CLL4, and the ARCTIC and ADMIRE clinical trials, respectively. This facilitated the selection of patient samples with short telomeres (TL<3.81Kb) used for the following datasets.

A total of 276 CLL patient samples, the majority with short TL, were identified and screened with the 5p8:17p6:XpYpM:16p1:21q1 telomere fusion assay to study telomere dysfunction and fusion in CLL, particularly involving the 5p telomere (**Table 3.4**).

Table 3.4. CLL patient samples selected to study telomere dysfunction and fusion

SOURCE	TL (0-3.81Kb]	TL (3.81-5.59Kb]	TOTAL
UHW	26	7	33
LRF CLL4	39	-	39
ARCTIC and ADMIRE	204	-	204
TOTAL	269	7	276

3.4.6.1 CLL patient samples exhibit different frequencies of telomere fusions, including the 5p telomere

For high-throughput fusion analysis of 276 CLL patient samples with the 5p8:17p6:XpYpM:16p1:21q1 fusion assay, Southern blots were hybridised with the 5p8F probe followed by the 17p6+XpYpOG and 16p+21q probes in pairs. Sequence alignment of the probes to be paired indicated that there was no complementarity between them.

Frequency of fusions (number of fusions per diploid genome) was calculated per each patient sample as stated in **section 3.4.4.4**. For 10 fusion PCR reactions using 100ng gDNA per reaction, the total number of telomere fusion events detected equivalent to its frequency of fusions is indicated in **Supplementary Table 8**.

CLL patient samples with varied frequencies of telomere fusions were identified. **Figure 3.20A** shows an overview of the number of telomere fusions detected from a panel of 7 CLL patient samples with different TLs. The frequency of telomere fusions detected for each patient sample in the cohort was calculated and summarised in **Supplementary Table 8** and **Figure 3.20BC**.

From the total cohort, telomere fusions were detected in 71.74% (198/276) of the samples. These were arbitrarily classified in 51.81% (143/276) with a low frequency (6.00×10^{-6} - 1.80×10^{-5}), 14.49% (40/276) with a medium frequency (2.40 - 3.06×10^{-5}) and 5.43% (15/276) with a high frequency of fusions ($>4.20 \times 10^{-5}$) (**Figure 3.20BC**).

A similar trend was observed for each independent cohort; however, more telomere fusions were detected from the UHW samples compared with the LRF CLL4 and the ARCTIC and ADMIRE samples (**Figure 3.20C**). There is the possibility that better fusion reactions are obtained from fresh samples and the fusions frequency relate to the quality of the sample since UHW were prepared from fresh samples in contrast with LRF CLL4 and the ARCTIC and ADMIRE samples that were cryopreserved.

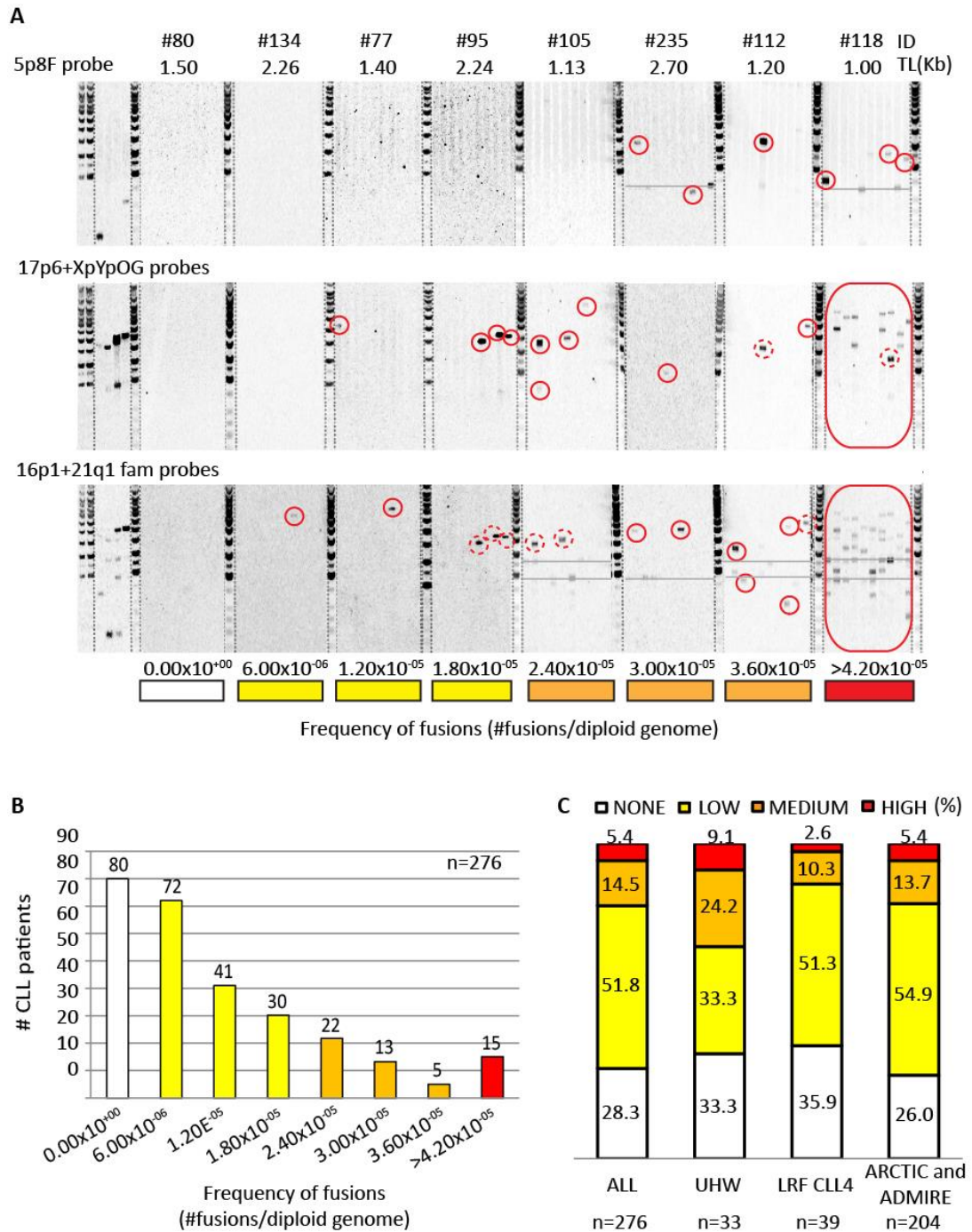


Figure 3.20. CLL patient samples exhibit different frequencies of telomere fusions.

(A) Panel of 8 CLL patient samples with different frequencies of fusions. Southern blots were hybridised with the 5p8F, 17p6+XpYp and 16p1+21q1 probes. Telomere fusion events are highlighted in red circles. To calculate the frequency of fusions dashed circles are not counted since they indicate events previously identified with different probes (potential inter-chromosomal fusions). (B) Frequency of telomere fusion in a panel of 276 CLL patient samples (10rxs, 100ng gDNA/rxs). (C) Distribution depending on cohort of origin of the sample: ALL (n=276), UHW (n=33), LRF CLL4 (n=39) and ARCTIC and ADMIRE (n=204).

3.4.6.2 The 5p telomere undergoes fusion in CLL

From the total cohort of 276 CLL patient samples, the first 33 samples (15 from the UHW and 18 from the LRF CLL4 clinical trial) were screened with the radiolabelled probes combined as follows: 5p+17p, XpYp+16p and 21q. However, the remaining 243 patient samples were screened using the 5p probe on its own, followed by the combined 17p+XpYp probes, and the 16p+21q probes. The second approach allowed the identification of potential novel 5p telomere fusion events (**Figure 3.20A**). In addition, the frequency of 5p to 17p+XpYp and 16p+21q potential telomere fusion events on the 72.84% (177/243) of patient samples for which fusions were detected was compared.

Potential 5p fusion events were identified in 22.60% (40/177) of the CLL samples for which telomere fusions were detected (**Supplementary Table 9**). **Figure 3.21A** illustrates an example of a 5p-XpYp fusion event in a CLL patient sample. This was similar to the proportion of CLL patient samples for which 17p+XpYp fusions were detected 28.25% (50/177), but not to the 16p+21q families (91.54%; 162/177) that was more than 3 times higher compared to the individual telomere probes (**Figure 3.21B**). However, once the frequencies are adjusted for the number of chromosomes detected by the probes, 5p telomere fusions are not rare compared to fusions involving the 17p and XpYp telomeres and the 16p and 21q families.

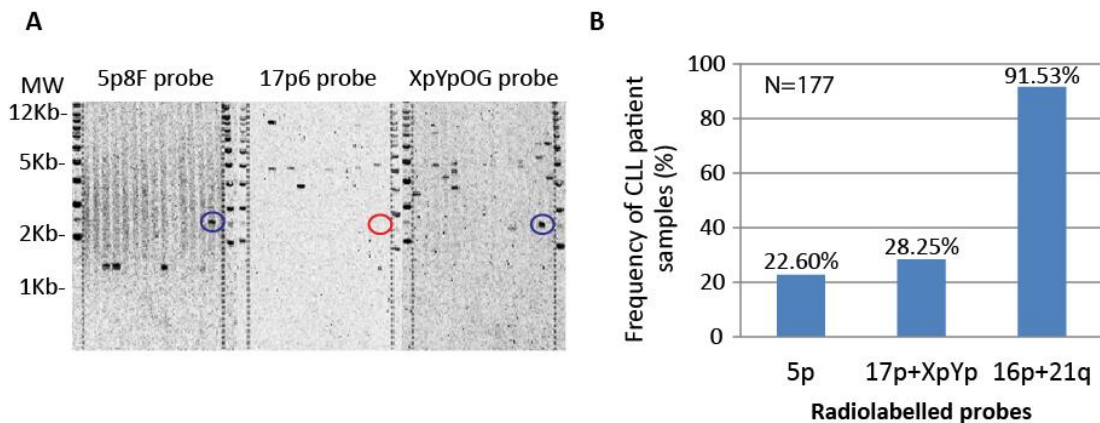


Figure 3.21. Detection of 5p fusions in CLL patient samples.

(A) Example of a 5p-XpYp telomere fusion event in a CLL patient sample. Southern blots were hybridised with the 5p8F, 17p6 and XpYp, respectively. Blue circles highlight the fusion event and the red circle indicates the absence of the event. **(B)** Proportion of CLL patient samples for which 5p, 17p+XpYp and/or 16p+21q telomere fusions were detected.

3.4.6.3 The frequency of fusions may provide prognostic information in CLL patients with short telomeres

Regression analysis was used to predict the relationship among the frequency of fusions and the XpYp TL for the 276 CLL patient samples from the UHW, LRF CLL4 and the ARCTIC and ADMIRE clinical trials. There was no linear association between shorter TL with an increase of frequency of fusions ($r^2 < 0.001$; $p = 0.876$). Similar results were observed from the ARCTIC and ADMIRE data on its own ($n = 204$; $r^2 < 0.0004$; $P = 0.784$) (**Figure 3.22A**).

At the time this project was undertaken, clinical data for 165/204 CLL patient samples from the ARCTIC and ADMIRE clinical trials was available (Howard *et al.*, 2017, Munir *et al.*, 2017). The next step was to investigate whether there was an association between cell proliferation and the frequency of fusions. Since the Lymphocyte Doubling Time (LDT) was not available, the Absolute Lymphocyte Count (ALC) was used; however, no significant correlation was observed ($r^2 = 0.0016$; $p = 0.609$) (**Figure 3.22B**).

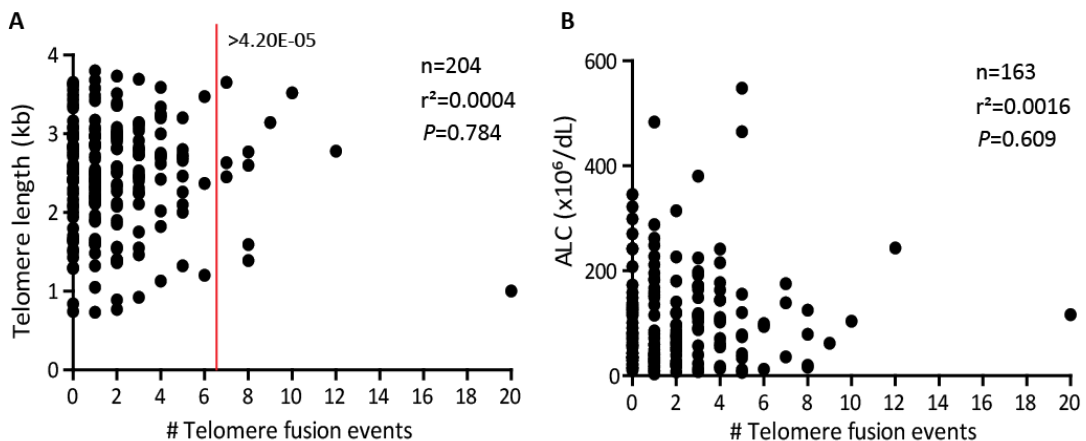


Figure 3.22. Telomere length, frequency of fusions and ALC.

(A) Association between the TL and the number of fusion events. **(B)** Association between the absolute lymphocyte count (ALC) and the number of fusion events.

A common CLL prognostic marker is the immunoglobulin heavy locus variable region (*IGHV*) gene mutational status since it identifies patients with worse (unmutated) and better (mutated) prognosis (Hamblin *et al.*, 1999, Damle *et al.*, 1999). The TL is a powerful indicator of patient survival in CLL (Lin *et al.*, 2010, Lin *et al.*, 2014) and was predictive of response to treatment in CLL patients from the ARCTIC and ADMIRE clinical trials ($n = 276$), particularly in the subgroup with mutated *IGHV* (Norris *et al.*, manuscript in preparation).

To investigate the prognostic resolution of the frequency of fusions in the subset of CLL patients with short TL (TL<3.81Kb) from the ARCTIC and ADMIRE (n=165), progression-free survival (PFS) and overall survival (OS) was compared with that obtained from the *IGHV* status (n=150).

PFS and OS from CLL patients with short telomeres with unmutated *IGHV* gene did not differ from those with mutated *IGHV* (n=150) [Hazard ratio (HR)=0.817, 95% confidence interval (CI)=0.483-1.383, $p=0.5665$] and (HR=1.299, CI=0.584-2.886, $p=0.5213$), respectively (**Figure 3.23AB**).

Patient's PFS and OS was assessed based on the presence and absence of fusions (HR=0.7363, CI=0.648-1.846, $p=0.7363$ and HR=1.726, CI=0.785-3.798, $p=0.1748$, respectively) (**Figure 3.23CD**). PFS and OS were re-examined based on the frequency of fusions categories: none, low-medium and high. Striking differences were observed between the three groups: the median PFS was 35.92 months for CLL patients with a high frequency of fusions; in contrast with 53.91 months for patients with med-low frequency and 51.68 months those with none (**Figure 3.23EF**).

PFS and OS based on the highest frequency of fusions or the absence of telomere fusions was further compared (n=53). Clear differences were observed between both subgroups and despite p -values improved, remained not significant (HR=1.390, CI=0.506-3.822, $p=0.5231$) for PFS and (HR=3.925, CI=0.797-19.33, $p=0.0927$) for OS.

Altogether, within a subgroup of CLL patients with short telomeres (TL<3.81Kb), patients with unmutated *IGHV* did not have significantly shorter PFS than patients with mutated *IGHV* (51.68 vs. 58.85 months). In contrast, patients with a high frequency of telomere fusions presented notably shorter PFS than patients with no fusions (35.92 vs. 51.68 months) despite significance was not achieved given the low number of patients (**Figure 3.23**).

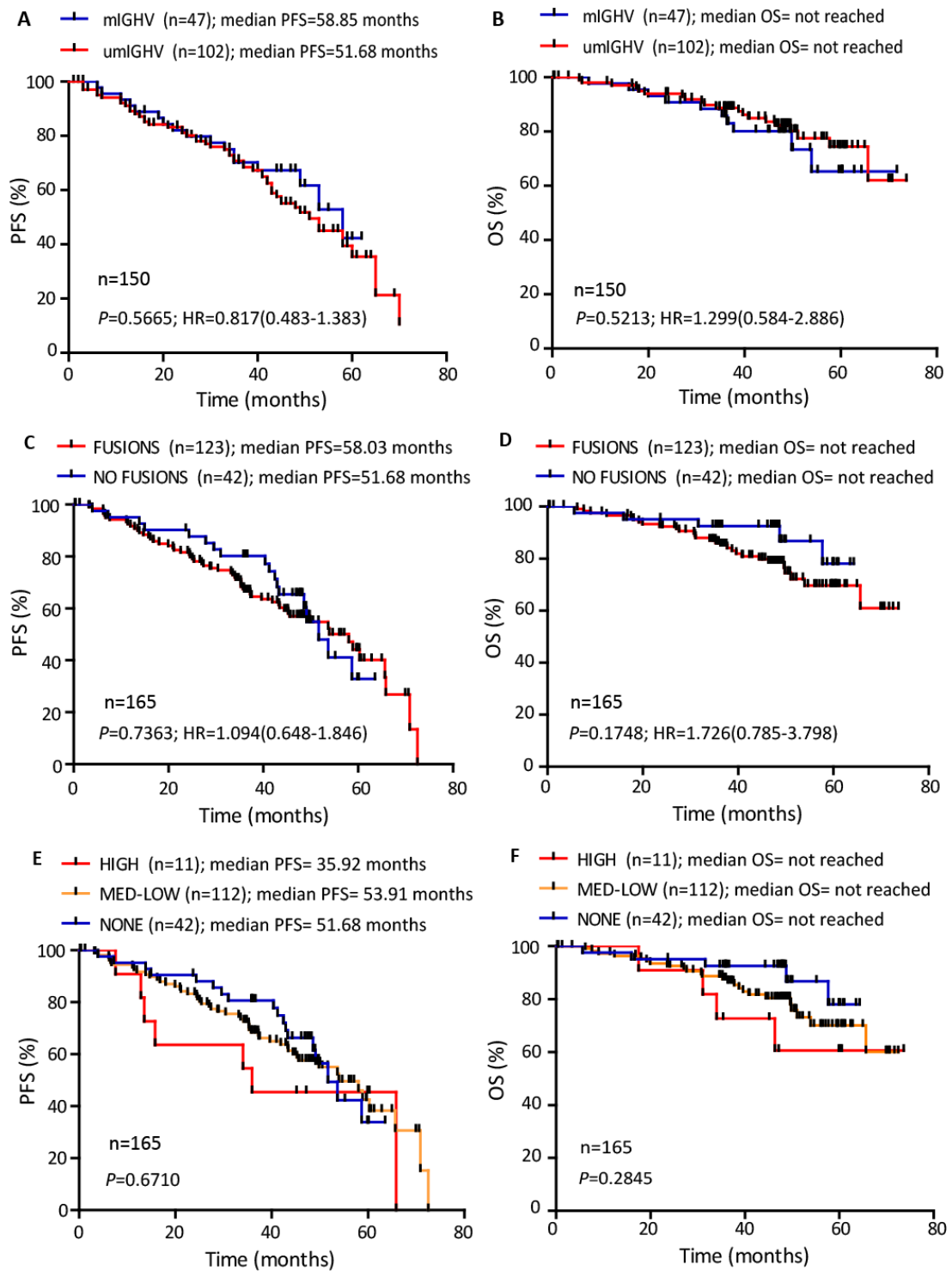


Figure 3.23. Frequency of telomere fusions as a prognostic marker in patients with short telomeres.

(A-F) Kaplan Meier curves of survival and the results obtained after Log-rank (Mantel-Cox) statistic categorising prognostic discrimination depending on (AB) *IGHV* gene status, (CD) presence or absence of fusions, and (EF) arbitrary categories of frequency of fusions for (ACE) PFS and (BDF) OS.

Clinical data from the ARCTIC and ADMIRE clinical trials were also available for CD38 cell marker expression (+,-), the *IGHV* gene status (mutated, unmutated), 11q23 and 17p cytogenetics (intact, deletion). To further investigate the impact of each independent CLL biomarker into telomere dysfunction and fusion, the proportion of CLL patients for which telomere fusions were detected, was compared for each prognostic marker. However, similar proportions of patients were observed within each category (**Figure 3.24A**).

To investigate the impact of the CLL drivers *NOTCH1*, *SF3B1*, *ATM*, *BIRC3* and *TP53* on telomere dysfunction and fusion, mutated (M) or unmutated (UM) gene status was compared to the proportion of patients within the LRF CLL4 clinical trial for which telomere fusions were detected. Surprisingly, when the CLL drivers were mutated (except for *BIRC3*), a higher proportion of patients with telomere fusions was detected. It remains possible that the lack of statistical significance is associated with the low number of patients in that cohort (**Figure 3.24B**).

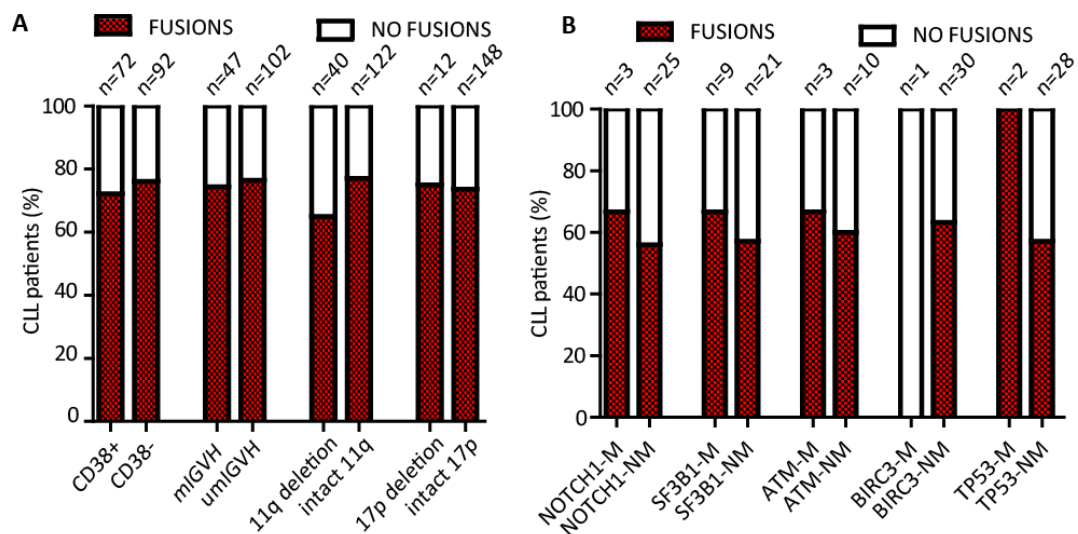


Figure 3.24. Frequency of fusions and other biomarkers.

(A) Biomarker status: CD38 expression, *IGHV* gene status, 11q23 deletion, 17p deletion. Chi-square test: CD38 ($p=0.3167$), VH ($p=0.0704$), 11q23 ($p=0.1313$) and 17p ($p=0.9185$). **(B)** CLL drivers: *NOTCH1*, *SF3B1*, *ATM*, *BIRC3* and *TP53*. Fisher's exact test (for small cell sizes <5): *NOTCH1* ($p=1.0000$), *SF3B1* ($p=0.7036$), *ATM* ($p=1.0000$), *BIRC3* ($p=0.3871$) and *TP53* ($p=0.5034$).

3.5 DISCUSSION

3.5.1 The 5p telomere length distributions are similar to the XpYp and 17p telomeres

The importance of STELA as a prognostic tool has previously been shown in a cohort of CLL patient samples measuring TL at the 17p and XpYp telomeres and demonstrating an association of shorter telomeres with decreased survival (Lin *et al.*, 2010, Lin *et al.*, 2014, Strefford *et al.*, 2015). In this chapter, 5p STELA has been successfully developed and validated to characterise the 5p telomere, providing a tool to specifically study this telomere that is proximal to the *hTERT* locus.

The measurement of the TL profiles in 57 CLL patient samples showed that 5p telomere was significantly longer than XpYp, but not to 17p (**Figure 3.5**). However, differences disappeared after TL measurement was corrected for the TVRs in 12 CLL patient samples (**Figure 3.6**). This data suggests that the differences in TL relate, at least in part, to the discrepant numbers of TVRs at the different chromosome ends, thereby explaining why the 5p telomere appears longer by standard STELA techniques. This study examined TCAGGG and TGAGGG, which together with TTGGGG, have been the most common TVRs identified in human telomeres (Allshire *et al.*, 1989, Baird, 1995, Lee *et al.*, 2014a, Baird *et al.*, 2003). However, a note of caution is due since in addition to the small sample size (n=12), the variety of at least 7 other variant telomeric repeat sequences (including GTAGGG, ATAGGG, CTAGGG, TTCGGG, TTTGGG, TAAGGG and TTAAGGG) have been identified in human telomeres in recent studies (human-derived cell lines: mortal, telomerase positive and ALT) (Letsolo *et al.*, 2010, Lee *et al.*, 2014a). Variants are found on the proximal regions and tend to vary on the TTA-sequence and be conserved in the GGG-sequence.

Together with the linear correlation of the distinct mean TL shown in **Figure 3.5**, an explanation for these results may be that distinct telomeres could have the same amount of canonical TTAGGG repeats but differences in length may arise from a variation in the number of TVR. In addition, these results are consistent with previous observations on the TVR composition of XpYp, 12q, 16p, 16q and 17p showing that each telomere has a unique distribution of TVRs (Baird, 1995, Baird *et al.*, 2003, Coleman *et al.*, 1999, Letsolo *et al.*, 2010).

The mechanism responsible for the generation of distinct TVRs remains unclear. According to Lee *et al.*, 2014 the most reasonable explanation for telomerase positive cells might be a

de novo synthesis of variant repeats misincorporated by telomerase. However, this explanation may not support our hypothesis as it would require telomerase presenting a different error rate for distinct telomeres. On the other hand, it has also been proposed that intra-allelic differences may occur from errors arising during replication including repeat deletion, duplication or mutation (Letsolo *et al.*, 2010).

The high variability of TVRs found at the proximal regions of human telomeres could indicate that these TVRs are not functional. In addition, some studies suggest that some TVRs could alter the formation of telomeric higher-order structures as well as decrease the binding affinity of the shelterin complex (Mendez-Bermudez *et al.*, 2009, Broccoli *et al.*, 1997a, Broccoli *et al.*, 1997b). Moreover, TVRs impact on accurate measurement of the pure TTAGGG repeat length, which used as a prognostic tool, can help to accurately identify patients at risk of disease progression (Lin *et al.*, 2010).

Overall, these results suggest a telomere-specific TVR pattern with same amount of canonical repeats. Further research should be undertaken to investigate whether the 5p telomere presents a higher proportion of TVRs. One possible way to assess this question would be to reamplify each STELA reaction for 5p and XpYp by NGS constructing a total characterisation of both alleles for all CLL patients. However, it may not extrapolate to all cases and it remains to be assessed in a bigger cohort group in CLL patient samples with longer telomeres. Particularly since it has been proposed that when telomerase is reactivated after cells escape a telomere-driven crisis, short telomeres become stabilised while longer telomeres carry on shortening (Ducray *et al.*, 1999).

In the literature there are contrasting views about chromosome specific TL. Some suggest that TL is proportional to the size of the chromosome arm while others propose that specific telomeres are shorter than others (Martens *et al.*, 1998, Deng *et al.*, 2004, Britt-Compton *et al.*, 2006, Wise *et al.*, 2009). In addition, substantial allelic variation has been reported by our group and others. There is evidence of bimodal distributions resulting from inheritance of maternal and paternal alleles by STELA and observations of distinct TL between chromosome homologs by Q-FISH analysis, which are suggested to be maintained throughout development (Baird *et al.*, 2003, Londono-Vallejo *et al.*, 2004, Graakjaer *et al.*, 2003). Evidence of different allelic distributions for 5p have also been reported by Q-FISH in HOSE cells (Deng *et al.*, 2004).

Moreover, CLL patients demonstrated variable TL profiles; some homogeneous, some heterogeneous and some bimodal. These results can be explained by the presence of a

dominant clonal B-cell growth resulting in homogeneous TL profiles, whereas heterogeneous or bimodal distributions likely result from the expansion of more than one major CLL-B cell clone. These results are consistent with previous observations in CLL patient samples (Lin *et al.*, 2010, Lin *et al.*, 2014).

3.5.2 The Fusion assay at the 5p chromosome end

To investigate dysfunction and fusion at the 5p telomere, a section of this chapter involved the adaptation of our group's telomere fusion assay (Capper *et al.*, 2007) to include the 5p telomere.

Similar to STELA, the primer design was the most critical step that would determine the specificity of the assay. In contrast to STELA, for the fusion assay the primers needed to be located several Kb away from the start of the telomere to be able to detect fusion amplicons that exhibit significant resection from the telomere (**Figure 3.7**). Our group has previously shown by Sanger sequencing of telomere amplicons, the presence of resection into the telomere, revealing fusion events lacking of TTAGGG repeats (Capper *et al.*, 2007).

The introduction of the 5p chromosome end into our group's fusion assay increased the number of fusion amplicons detected in the samples and allowed to detect 5p specific fusion events. The fusion frequency was similar at 5p compared to other ends, with the exception of HEK293 cell line that had a lower 5p frequency but a longer telomere. In addition, 5p fusion events may be under-represented. The CpG island within the 5p subtelomeric region (**Figure 3.7**) could have challenged the amplification of fusion events containing the GC-rich sequence (McDowell 1994). The lack of PCR product with the primer combination located at either side of the CpG island (**Figure 3.8**) further supported the idea that GC-rich sequence is refractory to long range PCR since it can form secondary structures that resist denaturation and impede primer annealing (McDowell *et al.*, 1998). Sequencing 5p fusion events will uncover whether fusions containing the CpG island can be detected.

3.5.4 The 5p telomere dynamics on HCT116^{DN-hTERT} before, during and after crisis are similar to the XpYp telomere.

Jones and colleagues showed that Ligase 3 (dependent A-NHEJ) is required to escape from telomere driven crisis, a critical event in the progression to malignancy. They also showed that large scale genomic rearrangements, which occasionally included the amplification of *hTERT* locus, were characteristic of clonal populations that escaped from crisis (Jones *et al.*, 2014). Two examples were HCT116 WT clone 11 and LIG3^{-/-mL3+nuc. LIG3} clone 15 (as shown in **Figure 3.15**) for which the 5p telomere was investigated in this study.

The comparison of the 5p to the XpYp TL revealed that the 5p telomere was significantly longer than XpYp, with a significantly linear correlation ($r^2=0.7047$). These results were consistent with what was shown for patient CLL-B cells before correcting for TVRs (**Figure 3.5**). As shown in **Figure 3.16**, both cell lines presented very short XpYp telomeres during crisis, and it is likely that the shorter telomeres are lost from STELA because they have been subjected to fusion, explaining the decreased number of telomeres detected. In these cases, using 5p STELA may be a more robust assay that would allow to study TL erosion during culture in cell lines with extremely short XpYp telomeres, or measure TL in patient samples.

Previous research has detected critically short 5p telomeres before crisis and showed evidence of 5p fusions with 19p and 20q, detected by Q-FISH and SKY in HOSE cells (Deng *et al.*, 2004). In addition, another study using FISH identified a 5p-18q fusion/duplication event involving the *hTERT* locus in all metaphases for HA1 cell line, suggesting some selective advantage. However, whether this event contributed to the reactivation of telomerase remains unknown (der-Sarkissian *et al.*, 2004). To study whether there was evidence of instability at the 5p telomere, the telomere fusion assay was performed for HCT116^{DN-hTERT} WT clone 11 and LIG3^{-/-mL3+nuc. LIG3} clone 15. According to previous research, fusions appear when very short telomeres are detected (Capper *et al.*, 2007). Potential 5p fusions were detected for both clones indicative of instability at this telomere prior to amplification of the 5p chromosome arm and reactivation of telomerase. Four very long (9-12.5Kb) 5p telomere fusion events (intra- and inter-chromosomal) were detected for HCT116 WT clone 11 in the last time point during crisis (day 66) (**Figure 3.19**). Previous research has shown that telomere sister chromatid fusion can cause gene amplification and deletion that drive carcinogenesis (O'Hagan *et al.*, 2002, Murnane, 2012). A plausible hypothesis agreeable with this would be 5p-5p intra-chromosomal fusions and BFB cycles

may provoke amplification of its chromosome arm which contains the *hTERT* locus, reactivating telomerase and driving malignant progression. This association has not been confirmed from this dataset but results indicate that there is instability at distinct telomeres, including 5p, during crisis.

There was no evidence for duplication or LOH at the 5p telomere after crisis and the 5p telomere displayed similar dynamics to those observed at XpYp. The *LIG3^{-/-}-mL3+nuc. LIG3* clone 15 presented with a very heterogeneous TL profile after day 153 (**Figure 3.16**). This may indicate the presence of distinct subclonal populations with different TL arising after a telomere-driven crisis. In addition, it remains possible that telomerase was not amplified in all subclones. Intra-tumour heterogeneity is a frequent source of relapse in cancer, particularly in CLL (Nowell, 1976, Landau *et al.*, 2014, Jacoby *et al.*, 2015), and telomere-driven crisis may provide a source for genomic recombination and selective pressure.

3.5.5 Telomere dysfunction and fusion in CLL, including the 5p telomere.

A cohort of 276 CLL patient samples with short telomeres, selected to be within the fusogenic range, was screened using the new optimised 5p8:17p6:XpYpM:16p1:21q telomere fusion assay. Telomere fusions were not detected in 29% of these patients, whilst different frequencies of telomere fusion events were detected in the remaining 71% (**Figure 3.20**). According to previous studies (Lin *et al.*, 2010, Lin *et al.*, 2014), these data demonstrate that there is significant telomere erosion in CLL that leads to telomere dysfunction and fusion during the progression of this disease.

One intriguing finding is that CLL patient samples with short telomeres exhibit different frequencies of fusions (**Figure 3.20**). There are several possible explanations for this result. It could be product of the limitation of the technique: not all telomeres are included; fusion junctions need to be within the location of the fusion primers and complex telomere fusion events may not be detected. There is also the possibility that fusions occurred earlier in the progression of the disease and are absent from the expanded surviving clones. Moreover, variation amongst patients could arise from faulty *TP53* or *ATM*, telomere capping proteins such as *TRF2*, *POT1* or *TPP1*, or DNA repair machinery (Artandi *et al.*, 2000, Chan and Blackburn, 2003, Denchi and de Lange, 2007, Arnoult and Karlseder, 2015, Guieze *et al.*, 2016). In future investigations it might be possible to assess these questions. To further confirm the presence or absence of fusions, Q-FISH could be used in addition to the fusion assay. Subsequently, sequencing known specific mutations on a panel of CLL patient

samples that presented different frequencies of telomere fusion events, would be able to clarify the mechanism by which some cells with short telomeres are more susceptible to telomere fusion than others.

The 5p8:17p6:XpYpM:16p1:21q1 fusion assay has also allowed the identification of CLL patient samples that exhibit the highest frequency of fusions. These samples were used to provide a detailed characterisation of telomere fusion in CLL using Next Generation sequencing, as detailed in the following chapter.

Another important finding is that fusions involving the 5p chromosome arm have been detected in patient CLL-B cells for the first time, indicating that this telomere can undergo recombination events that may contribute to genomic instability in CLL. The proportion of patients for which 5p fusions were detected was comparable to that obtained with the 17p and XpYp probes (**Figure 3.21**). Telomere-driven gene translocation and amplification can result in oncogene activation (O'Hagan *et al.*, 2002). Potential intra-allelic (sister chromatid) fusions involving 5p may provide insights into the mechanism of amplification or reactivation of the *hTERT* locus, particularly in CLL patients (Nagel *et al.*, 2010). In addition, detection of inter-chromosomal fusion events such as 5p-17p may be relevant to cancer progression since the 17p arm harbours the *TP53* tumour suppressor gene (Deng *et al.*, 2008). Further studies need to be undertaken to assess whether there is a direct association between 5p fusions with *hTERT* translocations.

Taken together, the addition of the 5p telomere was an improvement of our group's fusion assay as it detected an increased number of telomere fusion events and it has allowed the detection of potential 5p telomere fusion events in HEK293, HCT116 and patient CLL-B cells. Further research will be undertaken in the following chapter to characterise these 5p fusions events.

3.5.6 Telomere fusion frequency and prognosis in CLL patients with short telomeres

Short dysfunctional telomeres can fuse to other telomeres or dsDNA breaks causing genomic rearrangements that may drive cancer progression. Research in our group has shown that the upper TL threshold at which fusions occur (3.81Kb) particularly the fusogenic mean TL (2.26Kb) is highly prognostic in CLL patients (Lin *et al.*, 2010, Lin *et al.*, 2014, Strefford *et al.*, 2015). Therefore, this project aimed to investigate whether the

frequency of telomere fusions could further define prognosis of patients with short telomeres, and whether there was any association with other CLL biomarkers.

An association between the TL and the frequency of fusions was not found, suggesting there was no correlation between shorter TL and higher the frequency of fusions (**Figure 3.22A**). These results may indicate that when it comes to TL, below the threshold at which telomeres become dysfunctional they might be equally likely to undergo fusion.

To investigate whether patients with a higher proliferation rate of CLL-B cells have increased frequency of fusions the absolute lymphocyte count (ALC) was compared, as the lymphocyte doubling time (LDT) was not available, but no association was found (**Figure 3.22B**). Analysis of the CLL patients from the ARCTIC and ADMIRE cohorts has shown that TL is predictive of response to the treatment (Norris *et al.*, manuscript in preparation). To find out whether prognosis of CLL patient samples with short telomeres could be further stratified based on the presence of telomere fusion event, survival was studied based on different parameters. Progression-free survival (PFS) and overall survival (OS) was assessed in the ARCTIC and ADMIRE clinical trial depending on the *IGHV* locus (n=150) and the frequency of fusions (n=165). Prognosis based on the *IGHV* gene status, formerly used to identify patients likely to have poor survival, failed to stratify patients with short telomeres (**Figure 3.23AB**). Neither the presence or absence of telomere fusion, nor the fusion frequency (categories: high, medium-low and none), was able to stratify patients for PFS and OS. Whilst this was not statistically significant, it is important to note that a subgroup with the highest frequency of fusions ($>4.20E^{-05}$) presented 15 month shorter PFS and shorter OS than the subgroup with absence of fusion events (**Figure 3.23EF**). The trend observed proposes that in addition to the TL, it remains possible that the frequency of fusions may further stratify patients' prognosis for PFS and OS. Further research with an increased cohort size and a more sensitive technique should be undertaken.

Biomarkers including CD38 expression, *IGHV* gene mutation status and genomic aberrations such as 11q (*ATM*) and 17p (*TP53*) deletions, are valuable prognostic markers in CLL (Krober 2002). From the ARCTIC and ADMIRE CLL cohort, no correlations were found between the prognostic markers and the presence or absence of fusions, or the frequency (**Figure 3.24A**). These were unexpected results, particularly for the poor prognostic markers 11q (*ATM*) and 17p (*TP53*) deletion, for which inactivation of the *ATM* or *TP53* genes is associated with telomere dysfunction and fusions (Pettitt *et al.*, 2001, Oscier *et al.*, 2010, Guieze *et al.*, 2016). This inconsistency may be due to a deficient capture of all telomere

instability present in the samples. Alternatively, fusions might have occurred earlier in the progression of the disease and are absent from the expanded surviving clones.

From the CLL patients enrolled in the LRF CLL4 clinical trial (Strefford *et al.*, 2015), 39 with short telomeres were studied. It is interesting to note that the only 2 patients that presented *TP53* mutations from the 29 screened belong to the group that presented telomere fusion events. No significant associations were found between the presence of mutations in CLL drivers (*NOTCH1*, *SF3B1*, *ATM*, *BIRC3* and *TP53*) and the presence of fusions detected for each patient (**Figure 3.24B**). One possible explanation may be a limited dataset as most patients within the cohort were not screened for specific mutations.

These findings provide important insights into telomere erosion and dysfunction and the impact on patient survival in a subgroup presenting short telomeres. Despite the fusion frequency failed to significantly stratify patients with short TL, shorter survival was observed amongst patients with a high frequency of fusion compared with patients without or with lower frequencies of fusions. It remains to be further clarified in bigger cohorts, using a more sensitive technique, whether the frequency of fusions carries significant prognostic information and its association with other biomarkers.

3.5.7 Summary

The novelty of this study has been the successful adaptation of 5p STELA, TVR-PCR and the telomere fusion assay at 5p, that has allowed the characterisation of the 5p telomere (proximal to the *hTERT* locus) in different cell lines as well as in patient CLL-B cells.

The first section of this chapter consisted of the development of 5p STELA which can be used to investigate telomere erosion and importantly as a prognostic tool. Moreover, results showed that the 5p telomere was significantly longer than the XpYp but not the 17p. However, TVR analysis suggested that distinct telomeres may have the same amount of canonical TTAGGG repeats but differences in length may arise from the variation of TVRs.

The successful adaptation of the 5p8:17p6:XpYpM:16p1:21q1 telomere fusion assay has enabled, for the first time, the identification of novel 5p fusions in patient CLL-B cells. Potential fusion events involving 5p have also been detected in different cell lines, particularly HCT116 cells prior to escaping a telomere-driven crisis by amplification and reactivation of telomerase. These results demonstrate that there is telomere dysfunction and fusion at the 5p telomere. Further research is required to investigate whether it is a possible mechanism of translocation or amplification of the *hTERT* locus.

The analysis of the clinical data failed to identify any significant association between the fusion frequency and patient survival. However, a poorer prognosis was observed among a subgroup of CLL patients with short TL and the highest frequency of telomere fusions. Further research should be undertaken, using a more sensitive technique and with an increased cohort size, to investigate whether, on top of the TL, the frequency of fusions may be useful for a more accurate disease prognosis.

The newly adapted fusion assay has enabled the identification of CLL patient samples with the highest frequency of fusions which has permitted the continuation of the Ph.D. project allowing the high-throughput characterisation of telomere fusion events, including 5p, using a specialised NGS approach that will be discussed in the following chapter.

CHAPTER 4:

INVESTIGATING THE IMPACT OF TELOMERE DYSFUNCTION ON THE CHRONIC LYMPHOCYTIC LEUKAEMIA GENOME

4.1 ABSTRACT

Short dysfunctional telomeres are capable of fusion with one another or non-telomeric DSBs. The resulting cycles of telomere breakage-fusion-bridge can result in genomic rearrangements including amplification, deletion, translocations, chromothripsis and tetraploidisation. High levels of genomic instability are detrimental for cell viability; however, under certain conditions it may provide a selective advantage to the cell that can drive clonal evolution and malignant progression.

Telomeres, within the length ranges at which telomere fusions are detected, provide a prognostic marker for CLL. In chapter 3, telomere fusions were detected in 71.7% of 276 CLL patient samples with short telomeres ($TL < 3.81$ Kb). The aim of chapter 4 was to understand the role that telomere dysfunction plays in driving the evolution of the CLL cancer genome. A large-scale analysis was undertaken on single-molecule amplified telomere fusion events ($n=914$), obtained from 9 CLL patient samples that exhibited the highest fusion frequencies ($>4.20 \times 10^{-5}$ per diploid genome). The sequence of the fusion events was characterised using high-throughput paired-end sequencing.

This study examined head-to-head intra- and inter-chromosomal telomere fusions and identified non-telomeric genomic loci that fuse with dysfunctional telomeres. These loci included the ancestral telomere at Chr2q13, mitochondrial DNA and protein coding genes expressed in CLL-B lymphocytes and other oncogenes. Distinct microhomology usage at the fusion junctions was revealed, potentially implicating C-NHEJ, A-NHEJ and SSA or other DNA repair mechanisms mediating specific types of telomere fusion events. In addition, 5p sister chromatid fusion events were detected, which could lead to chromosomal instability that may impact on telomerase expression. Together, these data provide information on the role that telomere dysfunction and fusion plays in shaping the CLL genome and driving tumour heterogeneity.

4.2 INTRODUCTION

Telomeres play an important role in maintaining a stable genome. Dysfunctional telomeres are recognised as DSBs that activate the DNA damage response (DDR) and lead to cellular senescence or apoptosis. When cell cycle checkpoints are deficient, cells may enter a telomere-driven crisis state, where the inappropriate repair of telomeres can initiate cycles of chromosomal BFB (McClintock, 1941, Counter *et al.*, 1992, Ma *et al.*, 1993).

During crisis, cells accumulate genomic instability. Telomere fusions can result in chromosome gains and losses (aneuploidy), and large-scale genomic rearrangements like nonreciprocal translocations (Riboni *et al.*, 1997, Artandi *et al.*, 2000). Persistent chromatin bridges might escalate to cytokinesis failure and result in tetraploidy, whole-genome duplication (Pampalona *et al.*, 2012). The fusion of sister chromatids can result in deletions and gene amplification, a source of oncogene activation (Murnane, 2006). More recently discovered, telomere dysfunction and cycles of breakage-fusion-bridge (BFB) can also lead to chromothripsis and kataegis in several cancer cells (Maciejowski *et al.*, 2015, Ernst *et al.*, 2016). Genomic instability can be deleterious for the cell and result in cell death. However, it can provide some survival advantage resulting in the escape of a malignant clone from crisis and possibly drive malignant transformation (Artandi *et al.*, 2000, Davoli and de Lange, 2012, Davoli *et al.*, 2010).

The TL range at which telomeres become dysfunctional has been used as a prognostic tool in several malignancies, including CLL, breast cancer, Multiple Myeloma and Myelodysplasia (Lin *et al.*, 2010, Lin *et al.*, 2014, Simpson *et al.*, 2015, Hyatt *et al.*, 2017, Williams *et al.*, 2017). Several recurrent chromosomal alterations are associated with CLL patient prognosis, including trisomy 12, a gain of the entire chromosome, and the loss of genetic information with deletion of 11q22-23, 13q14 and 17p (Dohner *et al.*, 1997, Dohner *et al.*, 2000, Tsimberidou *et al.*, 2009). Additional alterations comprise of 2p or 8q gain, 6q or 8p loss, and other less frequent aberrations (Salaverria *et al.*, 2015). The presence of short dysfunctional telomeres and telomere fusion in CLL B-cells is associated with large-scale genomic rearrangements (Lin *et al.*, 2010). Moreover, this is detected in early-stage patients prior to clinical progression. Taken together, these data are consistent with the view that CLL B-cell cells can undergo a telomere-driven crisis during malignant progression, that this occurs early and can confer both a poorer prognosis and response to treatment (Lin *et al.*, 2014). A variety of CLL driver point mutations have been identified, including *TP53* and *POT1* that are involved in telomere dysfunction (Wang, 2011, Puente *et*

al., 2011, Quesada *et al.*, 2012). Mutated *TP53* is associated with an increased frequency of telomere fusions (Davoli *et al.*, 2010, Hayashi *et al.*, 2015, Liddiard *et al.*, 2016), and *POT1* is a member of the shelterin complex responsible for protecting telomeres by sequestering the 3'-overhang (Denchi and de Lange, 2007). *POT1* mutations leave telomeres unprotected which can be recognised as DSBs and lead to telomere fusions in CLL (Ramsay *et al.*, 2013).

Chromothripsis and kataegis have also been identified amongst CLL patients (Stephens *et al.*, 2011, Edelmann *et al.*, 2012, Salaverria *et al.*, 2015, Ernst *et al.*, 2016). Chromothripsis is a novel form of genomic instability that was first described in a patient with an aggressive form of CLL (Stephens *et al.*, 2011). It is characterised by extensive chromosome fragmentation and reorganisation localised to specific regions, with copy number changes that can lead to loss of tumour suppressors or oncogene amplifications, and is found in several cancers (Forment *et al.*, 2012). Chromothripsis is associated with worse prognosis in CLL patients (Edelmann *et al.*, 2012) and it was detected in 21% (7/33) of cases with shorter telomeres from a cohort of CLL patient samples with comparable parameters (Ernst *et al.*, 2016). Chromothripsis of the 5p chromosome, including gains of the *hTERT* locus, has also been observed from 3/8 CLL patients with a chromothripsis-like pattern (Salaverria *et al.*, 2015).

Sanger sequencing of telomere fusion events has identified complex fusions including Xp/Yp-Xq/Yq events with the potential of creating ring chromosomes, in addition to telomeric fusions with genomic loci and the interstitial telomeric 2q13-14 locus (Letsolo *et al.*, 2010). The application of a high-resolution Next Generation Sequencing (NGS) has facilitated the high-throughput characterisation of telomere fusions (Liddiard *et al.*, 2016). Results from NGS have revealed that telomeres predominantly fuse with coding genomic sequence and has elucidated the role of C-NHEJ (LIG4) and A-NHEJ (LIG1/3) in mediating inter- and intra-chromosomal fusions, respectively, in addition to the mutational profile of each (Liddiard *et al.*, 2016).

Telomere dysfunction and fusion is associated with genomic instability and has been implicated in the progression of CLL (Lin *et al.*, 2010, Lin *et al.*, 2014). A large-scale characterisation of telomere fusions in CLL is required to further elucidate the impact of telomeres in the genome of patient CLL-B cells to reveal its effect on the pathogenesis of this disease.

4.3 AIMS OF THE PROJECT

The purpose of this chapter was to provide a detailed characterisation of telomere fusion in CLL. This was undertaken using specialised high-throughput NGS of single-molecule amplified telomere fusion events, in 9 CLL patient samples identified in chapter 3 that exhibited the highest frequency of fusions.

The aims of this chapter were as follows:

- To test whether the 5p8:17p6:XpYpM:16p1:21q1 fusion assay is scalable for NGS and able to identify and characterise authentic fusion events.
- To characterise 5p fusions in CLL patient samples, as *hTERT* is proximal to the 5p telomere.
- To identify distinct types of telomere fusions including head-to-head fusions, intra-chromosomal and inter-chromosomal events.
- To compare deletion at sister chromatids in intra-chromosomal fusion events.
- To explore the distinct Telomere Variant Repeats (TVRs) found at telomeres.
- To study the degree of microhomology at the fusion junctions to elucidate the DNA repair mechanisms involved.
- To interrogate which areas of the genome (if any in CLL) become incorporated into telomere fusions and what are the features of the surrounding DNA.
- To investigate whether there are multiple telomere fusion events disrupting the same genomic locus (genomic hotspots).
- To assess whether there are any key genes relevant to CLL or other oncogenic pathways disrupted by fusion events.

4.4 RESULTS

4.4.1 Specialised paired-end NGS of telomere fusion amplicons

4.4.1.1 9 CLL patient samples with the highest frequency of fusions selected for specialised paired-end Illumina HiSeq4000 PE100

In chapter 3, 276 CLL patient samples with short telomeres were screened using the telomere fusion assay. 9 CLL samples with the highest fusion frequency ($>4.20E^{-05}$ fusions per diploid genome) were selected for a specialised paired-end NGS of telomere fusion amplicons.

To select the optimal CLL-B cell genomic DNA concentration, the 5p8:17p6:XpYpM:16p1:21q1 fusion assay was conducted using 4 different DNA concentrations: 25, 50, 100 and 150ng (**Figure 4.1A**). 200 reactions using 100ng of DNA per reaction was selected in order to achieve the maximum number of genuine fusion events without increasing the likelihood of non-specific amplification or compromising the Illumina sequencing efficiency with the addition of excessive amounts of genomic input DNA.

For each sample, 200-300 telomere-fusion PCR reactions were performed depending on DNA availability, which represented 600-900 fusion molecules per sample. The reactions from each sample were pooled and purified using the Agencourt AMPure XP beads (Beckman Coulter). Following verification of the PCR reaction before and after purification (**Figure 4.1B**), samples were sent for paired-end Illumina HiSeq4000 PE100 amplicon sequencing at the Oxford Genomic Centre (DB17) and at BGI Hong Kong (DB59-66).

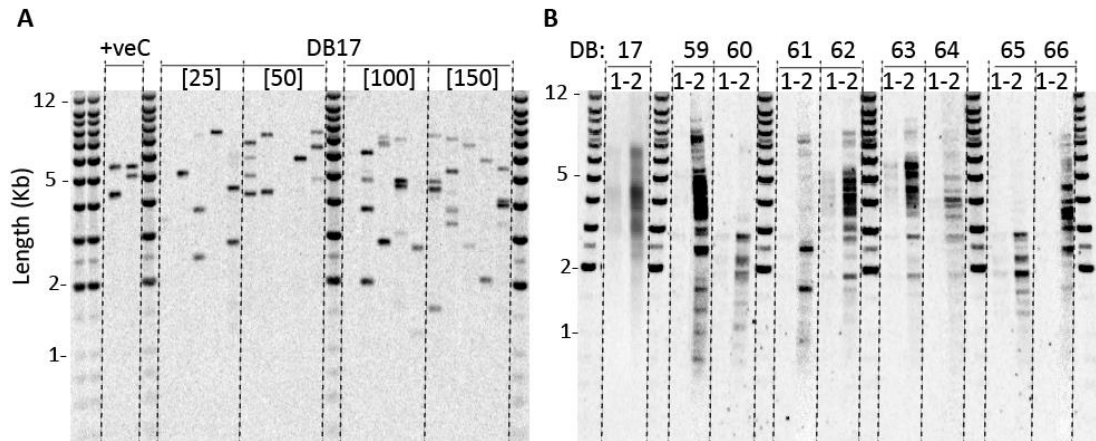


Figure 4.1. Preparation of 9 CLL patient samples for specialised NGS of telomere fusions.

(A) 5p8:17p6:XpYpM:16p1:21q1 fusion assay using 4 different DNA concentrations: 25, 50, 100 and 150ng of gDNA from patient DB17 CLL-B cells. Telomere fusions were detected using the XpYp probe. **(B)** Fusion PCR product of 200-300 reactions with a concentration of 100ng of gDNA/reaction. Before (1) and after (2) purification (10x more concentrated) of the PCR reactions using AMPure XP purification in preparation for Illumina paired-end sequencing of the fusion amplicons. Bands were detected with XpYp and 21q probes simultaneously.

Telomere length distribution, determined using STELA at the 5p, 17p and XpYp, revealed bimodal TL distributions (two distinct clusters of telomeres) in 67% (6/9) of CLL patient samples with the highest frequency of fusions (**Figure 4.2**). This frequency contrasts with the 4% (11/276) of bimodal CLL patient samples from the ARCTIC and ADMIRE cohorts (XpYp TL range: 0.80-7.49kb) as observed by Dr Norris using XpYp STELA (Norris *et al.*, manuscript in preparation).

It is unlikely that the bimodal TL distribution was a reflection of mixed cell populations rather than CLL-B cells since the UHW patient sample DB17 had been purified and contained nearly 100% CD19⁺ CLL-B cells. The remaining bimodal patient samples were from the ARCTIC and ADMIRE cohorts, therefore the CD19⁺ purity was not available. However, 4/5 samples displayed an absolute lymphocyte count (ALC) similar to, or above, the average 103×10^6 lymphocytes from the 276 CLL patient samples from the ARCTIC and ADMIRE cohorts: 163 (DB62), 19.6 (DB63), 243 (DB64), 79 (DB65) and 99 (DB66) $\times 10^6$ lymphocytes. Therefore, the vast majority of the cells analysed in this analysis were CLL B-cells and thus the bimodal distribution were a reflection of these cells and not normal cells present in these samples.

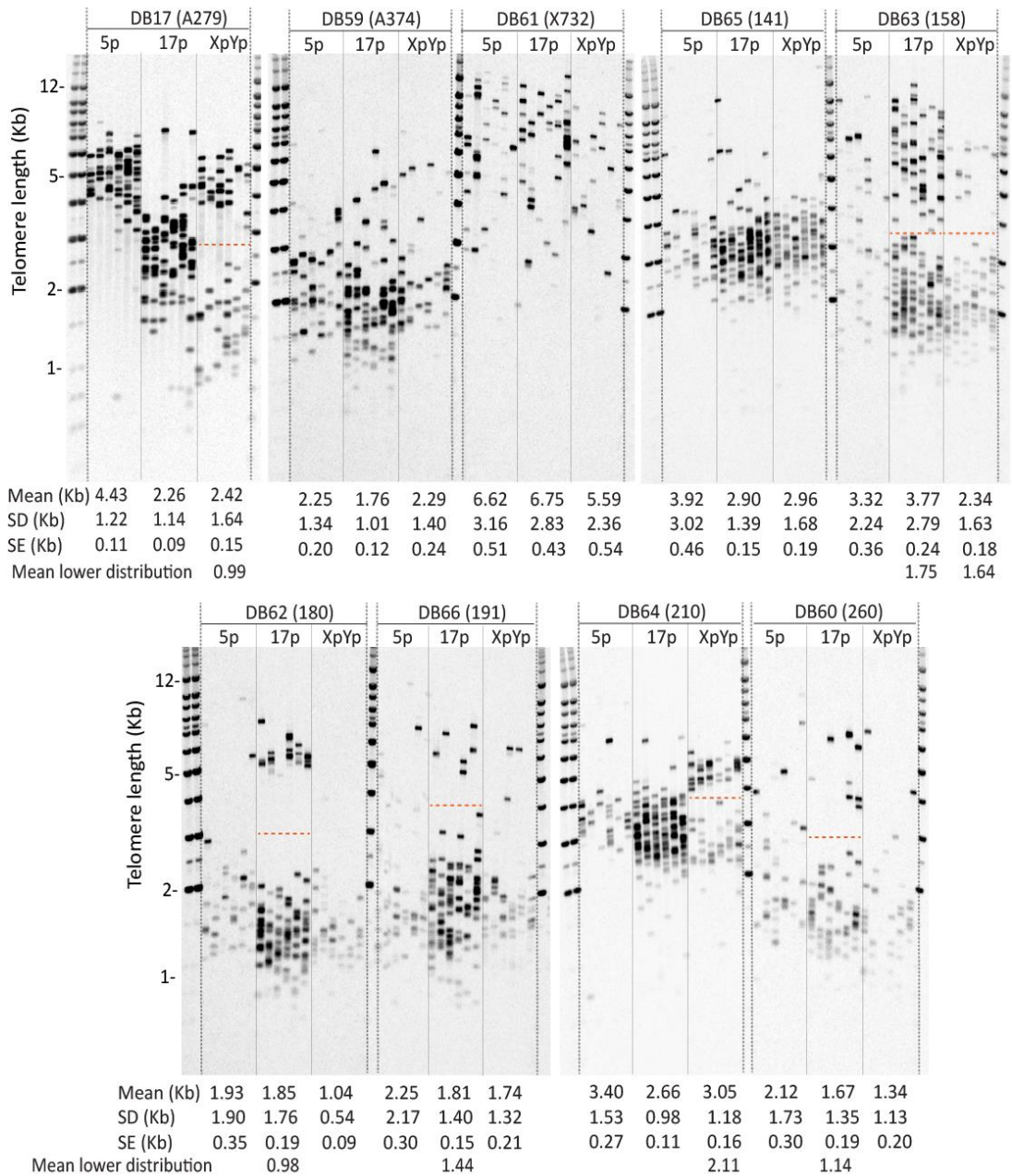


Figure 4.2. Bimodality identified in the CLL patient samples for NGS.

5p, 17p and XpYp STELA for the 9 CLL patient samples with the highest frequency of fusions selected for paired-end NGS of telomere fusion amplicons. Southern blot showing STELA for three different telomeres hybridised with the telomere probe. Mean TL, standard deviation (SD) and standard error (SE) indicated underneath. Bimodal distribution separated arbitrarily with a dotted red line and mean TL of lower distribution (Kb) indicated underneath. Clear bimodal distribution at 6/9 CLL patient samples: DB17, 60, 62, 63, 64 and 66. Heterogeneity or potential bimodality of DB59. NGS ID and corresponding CLL patient ID are as follows: DB17 (A279722), DB59 (A374068), DB60 (260), DB61 (X732750), DB62 (180), DB63 (158), DB64 (210), DB65 (141) and DB66 (191).

4.4.1.2 Mapping pipeline: intra- and inter-chromosomal analysis

Post-NGS, data for each sample was processed for mapping of intra- and inter-chromosomal telomere fusion events. Workflow for the analysis of telomere fusion amplicons is illustrated in **Figure 2.2**.

Sequence reads were trimmed to remove sequence tags and primers. Then, for the analysis of inter-chromosomal telomere fusion events, reads were mapped to the unmasked human hg19 reference sequence (modified according to the human subtelomeric sequence Stong *et al.*, 2014) and the 5p, 17p, XpYp, 16p and 21q subtelomeric sequences. Discordant read-pairs mapping to each chromosome end were selected for analysis of inter-chromosomal fusion events (**Figure 4.3**).

For intra-chromosomal fusion event analysis, read-pairs mapping to the 5p, 17 and XpYp telomeres only were filtered on orientation ($\rightarrow\rightarrow$) and only paired-reads both in forward (F) orientation indicative of sister chromatid head-to-head fusions were selected for analysis. The pipeline was developed by Dr Ashelford and the mapping and filtering of read-pairs was performed by Dr Cleal.

BAM and BED files were generated for visualisation of data spread along subtelomeric sequence in the IGV browser and read-pair coordinates were listed in txt files. Read-pairs with mapping quality (MQ) ≥ 0 were selected to exclude ambiguous or poor-quality mappings. Alignments obtained from the intra- and inter-chromosomal fusion analysis were further validated using BLAST (Altschul *et al.*, 1990) to human reference hg19 GRCh37 to certify unique and accurate mapping based on sequence identity and Expect (E) value significance (**section 4.3.1.4** and **4.3.1.5**). The identity is the proportion of the sequence searched that matches the other sequence in the database. The E-value describes the number of matches expected by chance doing a BLAST search in a particular size database. Events were classified according to **Table 4.1**.

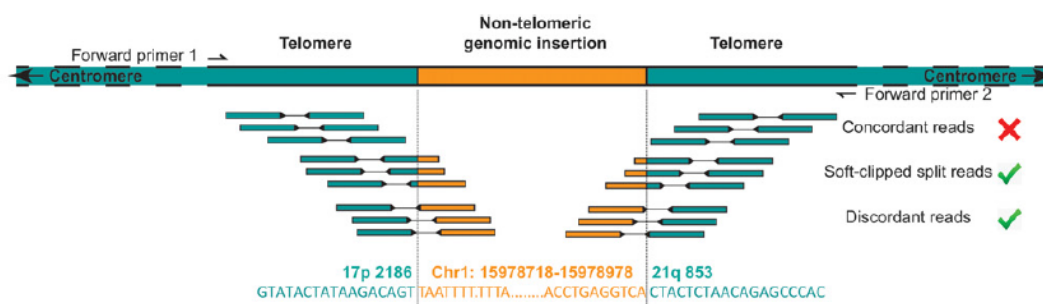


Figure 4.3. Characterisation of a telomere-genomic inter-chromosomal fusion event.

Cartoon representation of an inter-chromosomal telomere fusion with genomic locus sequenced by Illumina HiSeq4000 paired-end sequencing. Concordant read-pairs are those that map to the reference sequence with the expected size coverage and orientation while discordant read-pairs do not. Softclipped reads are partially discordant and carry information about the fusion junction. Example of a telomere-genomic 17p-Chr1p12-21q fusion event in crisis-stage MRC5^{HPV6E7} cells. Fusion PCR primer orientation indicated on the chromatids. Figure reproduced from Liddiard *et al.*, 2016.

4.4.1.3 Classification of events identified

All events recorded from the intra- and inter-chromosomal analysis were visualised on IGV and BLAST-alignment was performed to validate genuine telomere fusion events and exclude PCR or sequencing artefacts (4.4.1.4 and 4.4.1.5). Telomere fusions were further classified according to **Table 4.1** depending on the locations involved in the fusion event. A head-to-head telomere fusion was called telomere-telomere (00). When DNA sequence adjacent to the telomeric repeat was fused to telomeric repeat, allowing the identification of one of the chromosome end involved only, it was labelled subtelomere-telomere (0). The fusion of two subtelomeric sequences from the same or distinct chromosome arms was an intra-chromosomal (1) or inter-chromosomal (2T) event, respectively. Other types of inter-chromosomal fusion events involved the fusion of a telomere (including subtelomeric sequence or not) with the ancestral telomere at 2q13 (2A), genomic loci (2G) or mitochondrial DNA (2M). Complex events (2C) involved more than one of the previous groups. Telomere fusions between the 16p-16p and 21q-21q families could be either intra- or inter-chromosomal (1o2T) as most of the time it was not possible to determine which member of the family was involved. Excluded events were either unreliable (U) or present in other samples (“).

Table 4.1. Classification of type of event.

	Type	Type of event	Example
INCLUDED	00	Telomere-Telomere (head-to-head)	TTAGGG-CCCTAA
	0	Subtelomere-Telomere	5p-CCCTAA 5p-TTAGGG
	1	Intra-chromosomal fusion event (sister chromatid)	5p-5p
	2	Inter-chromosomal fusion event	
	2T	Inter: two distinct Telomeres *Including a potential ring chromosome formation	5p-Xp; Xp-Xq
	2A	Inter: telomere fusion with Ancestral telomere 2q13	5p-2q13
	2G	Inter: telomere fusion with Genomic loci	5p-Chr5p14.1
	2C	Inter: Complex telomere fusion event	5p-Chr13q14.2- Chr13q31.1-Tel
	2M	Inter: telomere fusion with Mitochondrial DNA)	5p-ChrM
	1o2T	Intra or Inter: fusion of the 16p or 21q family of telomeres	16p-16p; 21q-21q
EXCLUDED	U	Unreliable. Not confident reporting it: BLAST-alignment scores did not pass the threshold or primer fusion sequence was mapped (potential PCR artefact or background band)	
	"	Same event across all 9 CLL patient samples	

4.4.1.4 Validating intra-chromosomal fusion events

To further validate the read-pairs filtered from the intra-chromosomal mapping, all paired-reads and their coordinates were recorded in a txt file for manual curation and sequence verification for each CLL patient sample. Read-pairs were grouped into unique and distinct events based on location and fusion junction of the mapped and softclipped sequences. Each read-pair was visualised on IGV and aligned to the human genome (Ensembl BLAST) and 5p, 17p and XpYp subtelomeric sequences (Stong *et al.*, 2014). This allowed sequence verification and validation to certify the mapping corresponded to a unique and accurate location. A representative read-pair allowing maximum coverage for each event was further validated and scores resulting from BLAST-alignment (location, orientation, length, identity and E-value significance) together with the number of supporting reads were recorded.

From all 9 CLL patient samples, the intra-chromosomal analysis was performed independently for 5p, 17p and XpYp. It was not possible to study 16p or 21q sister

chromatid fusions since it was not possible to assess which member of the family of telomeres was involved. A total of $n=55$, $n=64$ and $n=69$ validated events for 5p, 17p and XpYp, respectively, were obtained for all patients that were subsequently combined and classified into the type of event according to **Table 4.1** (the summary of events is shown in **Figure 4.4** and **Supplementary Table 10**).

The intra-chromosomal analysis was used to detect sister chromatid fusion events (type 1). However, the vast majority of events filtered were potential ‘false’ positives. These consisted of subtelomere-telomere (type 0), most likely due to mapping the TTAGGG-sequence to the sister chromatid in same orientation ($\rightarrow\rightarrow$), and low-frequency telomere-telomere fusion events TTAGGG-CCCTAA (type 00). Telomeric inter-chromosomal fusion events (type 2T) were ‘false’ positives identified and were explained by the mapper’s incorrect subtelomeric alignment, corrected after BLAST (**Figure 4.4**). It remains unknown whether fusion events type 00 and 0 are intra- or inter-chromosomal since it was not possible to confirm the chromosome end involved.

Examples of each type of event (00, 0, 1 and 2T) for the 5p, 17p and XpYp from all 9 CLL patient samples are shown in **Figure 4.5**, **Figure 4.6** and **Figure 4.7**, respectively.

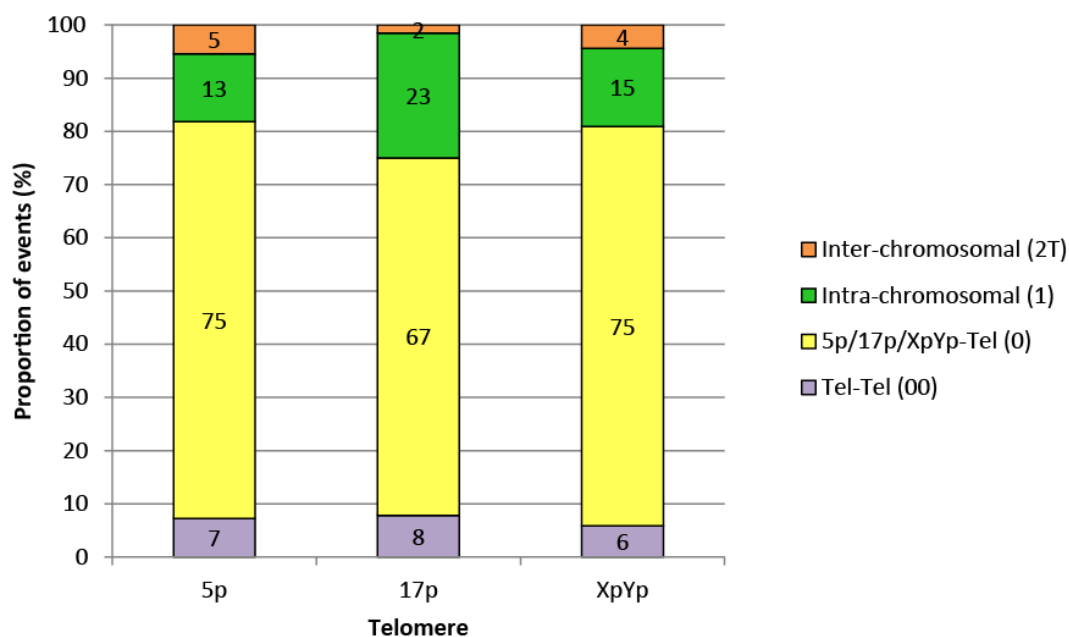


Figure 4.4. Summary of intra-chromosomal fusion analysis.

Summary of the type of telomere fusion event detected with the intra-chromosomal analysis for the 5p, 17p and XpYp telomeres. Statistical analysis with Two-way ANOVA and Bonferroni post-test. Telomere $p=1$, Type of event $p<0.0001$. There were no significant differences between 5p, 17p and XpYp for any type of event. Total 5p $n=58$, 17p $n=64$, XpYp $n=69$ events.

4.4.1.5 Validating inter-chromosomal fusion events

To further validate the inter-chromosomal events obtained for each patient sample, a list of coordinates recorded from the mapping pipeline containing ≥ 3 supporting read-pairs per event, and $MQ \geq 0$ was visualised in IGV. Mapped and softclipped sequence of each read-pair was verified and validated after BLAST-alignment to the human genome and the 5p, 17p, XpYp, 16p and 21q telomeric sequences (Stong *et al.*, 2014). Events were classified according to **Table 4.1**. An example of a telomeric inter-chromosomal fusion event (2T) is shown in **Figure 4.8**.

Following the inclusion of the 5p telomere into the fusion assay in chapter 3, an additional 56 events with < 3 supporting reads involving the 5p telomere were included. Of these, 4 were subtelomere-telomere (0), 1 was intra-chromosomal (1), 11 were telomeric inter-chromosomal (2T), 39 were 5p with genomic loci (2G, $MQ=60$) and 1 was a 5p-ChrM (2M, $MQ=29$).

Validation was based on stringent BLAST-alignment scores (identity and E-value significance) to ensure the read locations were unique. Proximity to potential locations of non-specific primer binding was verified after (low-stringency) BLAST of the fusion primers and noting potential binding locations.

From the 9 CLL patient samples a total of 728 fusion events were manually validated and verified. 46.29% were telomeric inter-chromosomal fusion events (2T), 30.91% were subtelomere-TTAGGG fusion events (0) and 10.30% were telomeric fusions with genomic loci (2G). Of the remaining 12.50% of events detected, 6.73% were telomeric intra- or inter-chromosomal fusion events (1o2T: 16p-16p or 21q-21q), 2.61% were sister chromatid fusion events (1), 1.51% were telomere-2q13 (2A), 0.69% were complex events (2C), 0.55% were telomere-mitochondrial (2M) and 0.41% were TTAGGG-CCCTAA (0) fusion events (**Figure 4.9**). Detailed summary of events detected for each CLL patient sample was shown in **Supplementary Table 11**.

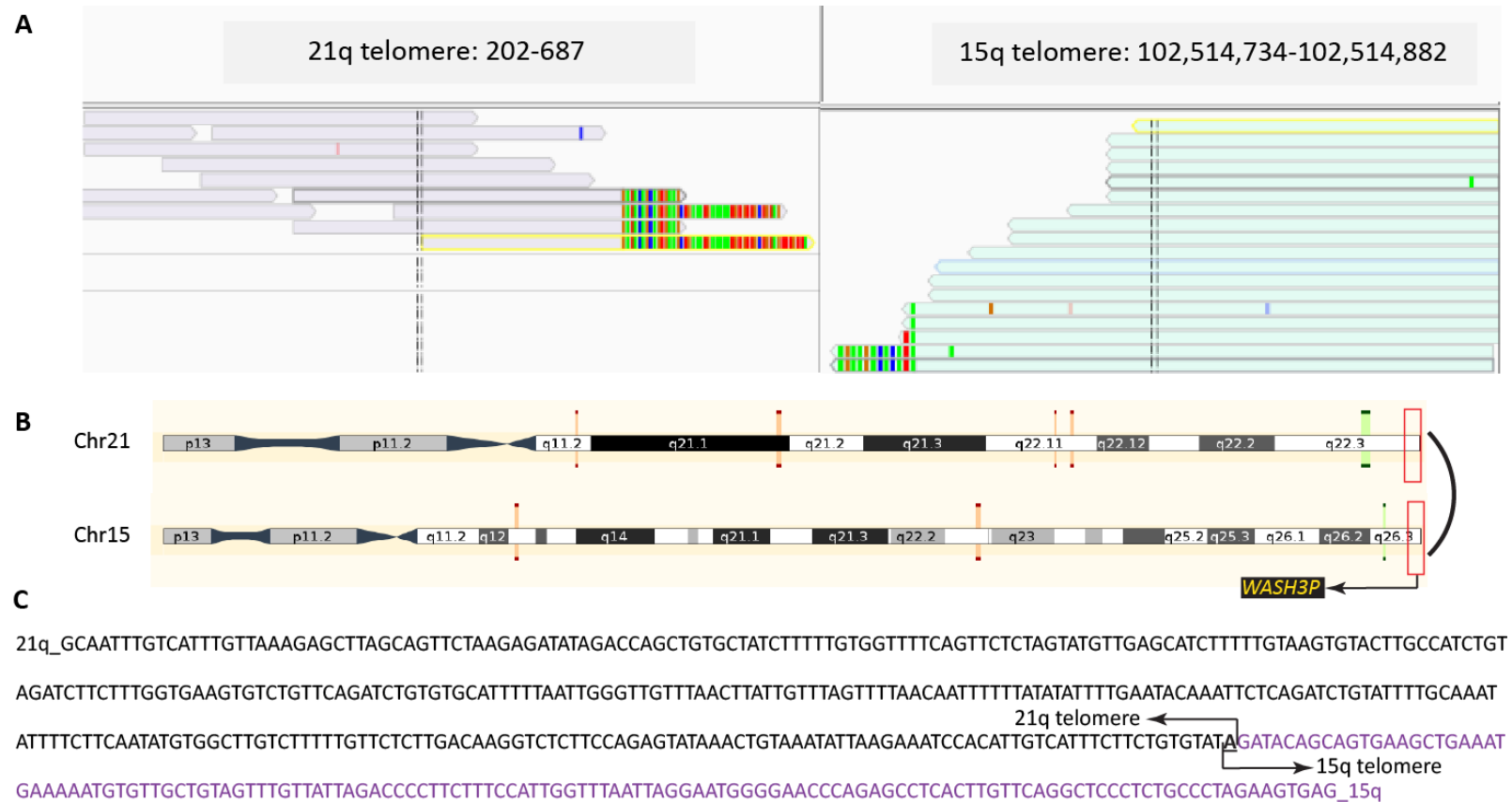


Figure 4.8. Telomeric inter-chromosomal fusion event.

(A) Paired-reads mapped for 21q and 15q (16pfam) telomeres together with breakpoints visualised in IGV. **(B)** Location of the reads for the telomere fusion event shown in Ensembl. Pseudogene identified *WASH3P* (*ENSG00000185596*) as indicated by Ensembl/Havana. **(C)** Sequence of the fusion event from DB17. The 21qfam telomeric sequence (unique to 1q or 21q in this case) represented in black, the 15q telomere in purple and microhomology at the junction is underlined.

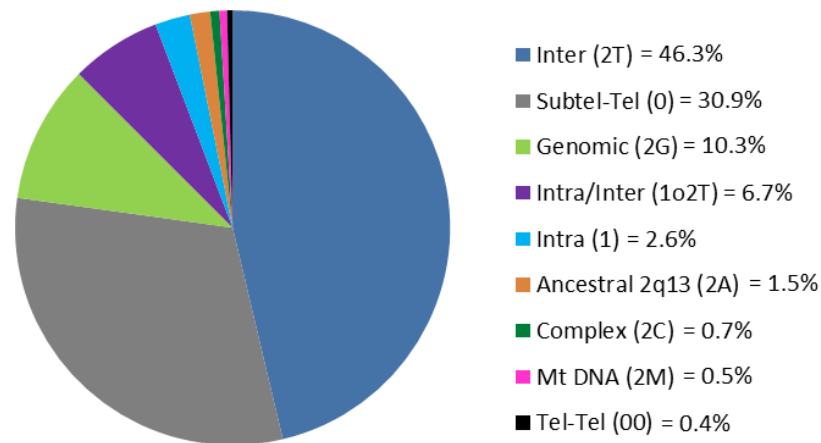


Figure 4.9. Summary of inter-chromosomal fusion analysis.

The proportion of type of telomere fusion event identified from the inter-chromosomal fusion analysis of 9 CLL patients.

From the inter-chromosomal analysis, a small proportion of sister chromatid fusion events were detected as a consequence of the mapper counting the 5p telomere, chr5 and 5p_Stong as distinct chromosomes. Surprisingly, the inter-chromosomal analysis identified an additional class of sister chromatid fusion events (n=5) that appeared to arise as a consequence of the insertion of subtelomeric DNA from the same telomere in reverse orientation. Such events have previously described in cell culture models (Capper *et al.*, 2007, Letsolo *et al.*, 2010) and may by arise from additional processing of the short telomeres prior ligation during BFB cycles (**Figure 4.10**).

Inversions were not identified among the sister chromatid fusion events detected with the intra-chromosomal analysis. It was not possible to assess for events with unmapped fusion junction (unclipped reads).



Figure 4.10. Sister chromatid fusion event detected from the inter-chromosomal analysis.

5p-5p intra-chromosomal fusion event potentially originated after a 5p-telomere fusion event (DB66). Microhomology at the fusion junction is highlighted in black and underlined. Different colours used to distinguish the distinct sections of the fusion event.

4.4.2 Signature of intra- and inter-chromosomal telomere fusion events in a panel of 9 CLL patient samples

Distinct types of telomere fusions were identified from the 9 CLL patient samples, but it remained to be assessed whether there was a distinct pattern of telomere fusions across the genome for each CLL patient sample. Data obtained from the intra- and inter-chromosomal analysis was compiled (**Supplementary Table 12**) and Circos plots displaying the linkages (Krzywinski *et al.*, 2009) were created from all validated and verified mapped reads as a composite (**Figure 4.11**).

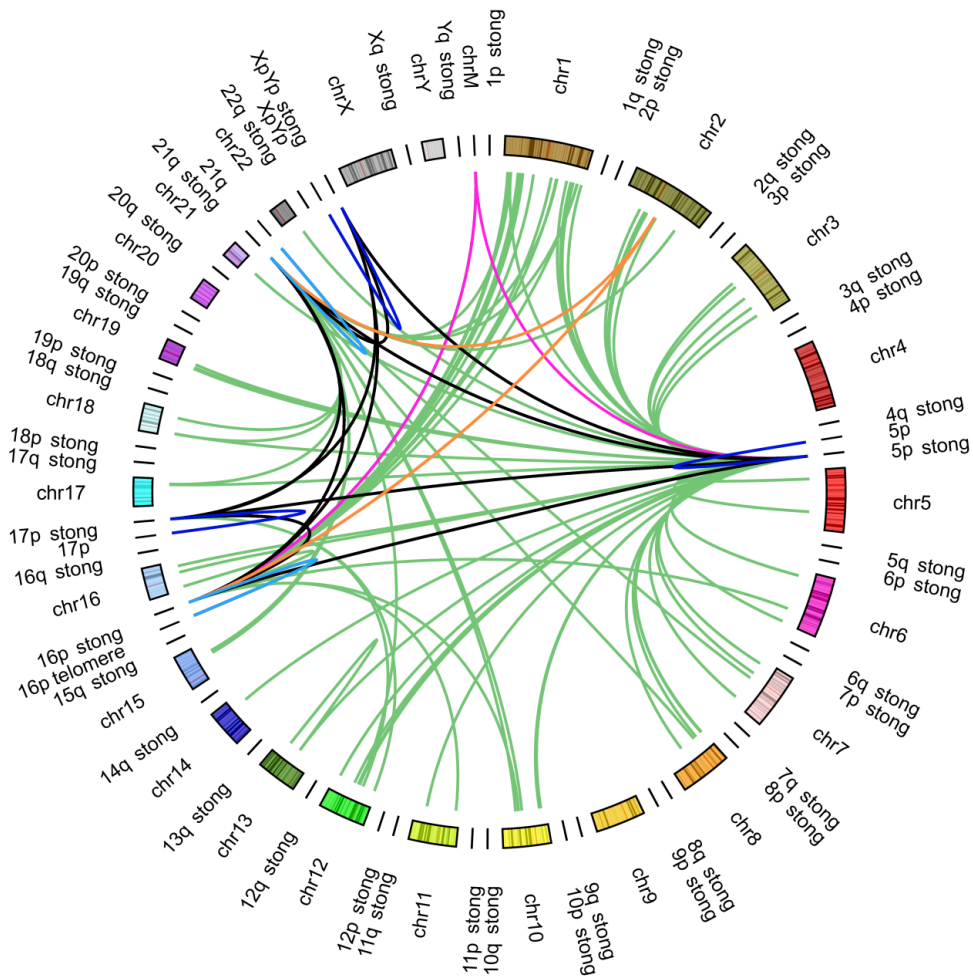


Figure 4.11. Characterisation of all telomere fusion events detected in 9 CLL patient samples.

Circos plot showing the validated results obtained from the inter- and intra-chromosomal telomere fusion analysis from 9 CLL patient samples. Circos plot with each chromosome and its telomeres (1p telomere, Chr1, 1q telomere) around the circle orientated clockwise. Colour code: telomere-telomere inter-chromosomal (black), telomere-telomere intra-chromosomal for 5p, 17p and XpYp (blue), inter- or intra-chromosomal for 16p and 21q families (light blue), and inter-chromosomal telomere-genomic (green), telomere-2q13 (orange) and telomere-ChrM (pink).

Independent plots were created for each CLL patient sample (**Figure 4.12** and **Figure 4.13**). Distinct signatures of telomere fusions were observed for the individual CLL patient samples. Some patients displayed simple patterns, characterised with intra- and/or inter-chromosomal head-to-head telomere fusions only (DB63 and DB60). Another patient mainly presented abundant genomic linkages, including the ancestral telomere at 2q13 and mitochondrial DNA (DB65). Finally, there were patients that exhibited complex signatures with a combination of most (DB17, DB59, DB61, DB62 and DB64) or all types of telomere fusion events (DB66).

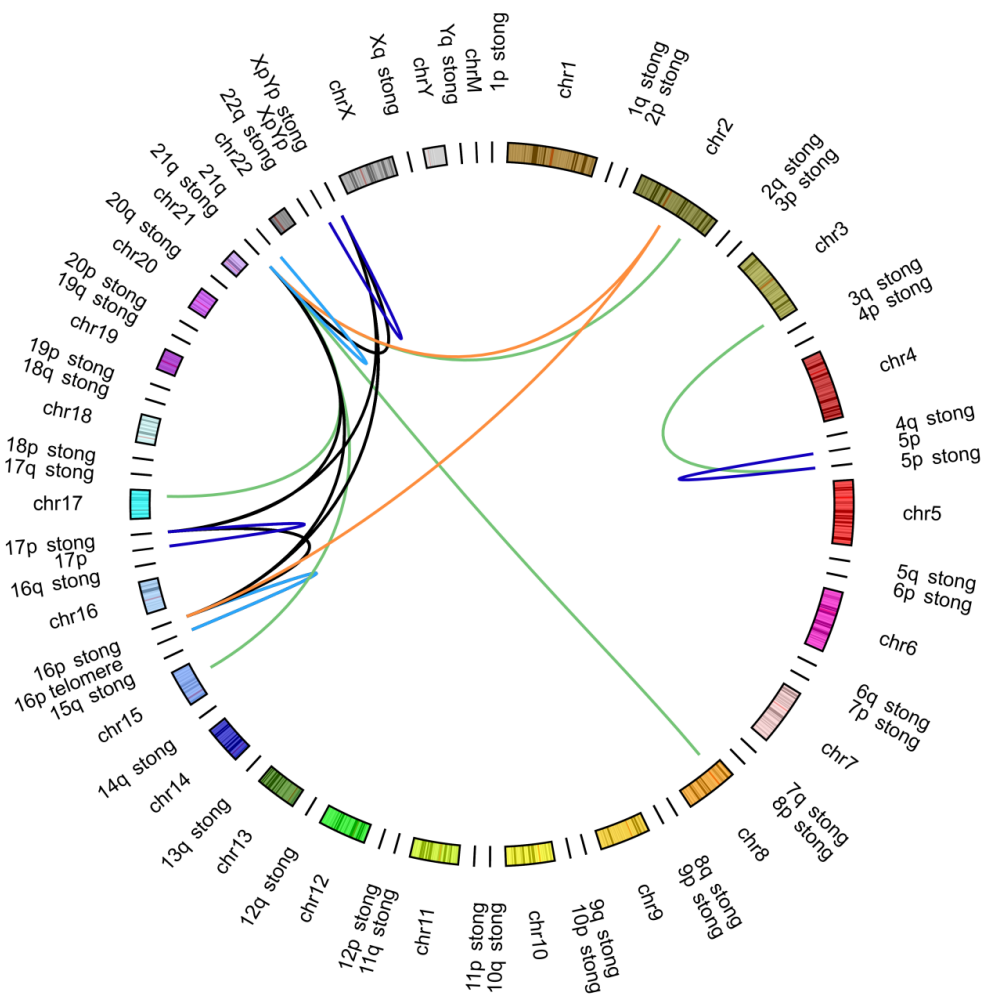


Figure 4.12. Characterisation of telomere fusions detected from CLL patient sample DB17.

Circos plot showing the validated results obtained from the inter- and intra-chromosomal telomere fusion analysis from CLL patient sample DB17. Circos plot with each chromosome and its telomeres (1p telomere, Chr1, 1q telomere) around the circle orientated clockwise. Colour coded as described in Figure 4.11.

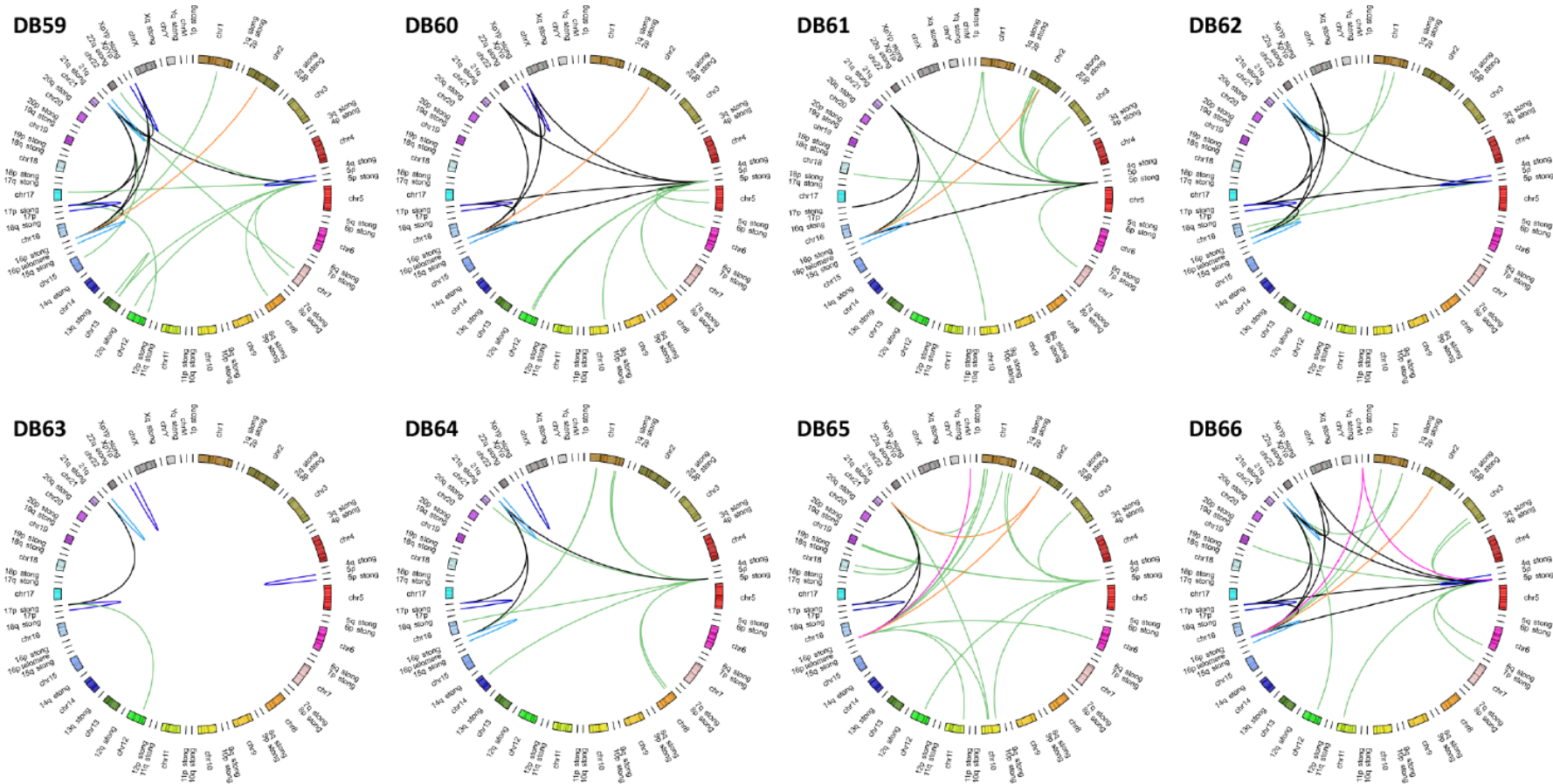


Figure 4.13. Characterisation of telomere fusions for 8 CLL patient samples.

Circos plot showing the validated results obtained from the inter- and intra-chromosomal telomere fusion analysis from 8 CLL patient samples (DB59-66). Circos plot with each chromosome and its telomeres (1p telomere, Chr1, 1q telomere) around the circle orientated clockwise. Colour coded as described in Figure 4.11.

Given the inclusion of the 5p telomere into the telomere-fusion assay it was reassuring to identify events involving this telomere. From 914 validated telomere fusion events detected from the panel of 9 CLL patient samples (**Table 4.2**), 19% (172 fusion events detected) involved the 5p telomere. These were distributed across the common classes of telomere fusion event that had been identified (**Figure 4.14**).

Table 4.2. Validated telomere fusion events identified from a panel of 9 CLL patient samples.

Fusion event	00	0	1	1o2T	2T	2A	2G	2C	2M
N=914	16	360	49	49	343	11	75	7	4
(%)	1.75	39.39	5.36	5.36	37.53	1.20	8.21	0.77	0.44

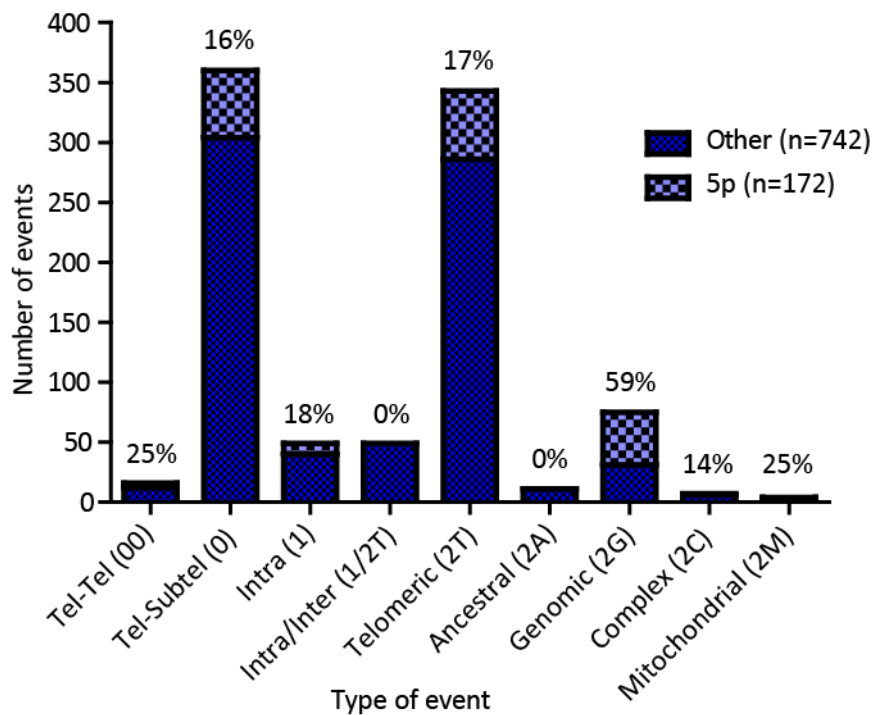


Figure 4.14. Proportion of 5p telomere fusion events.

Type of telomere fusion events detected with the 5p telomere (light blue chequered) and other telomeres: 17p, XpYp, 16p or 21q (dark blue). From a total of 914 events, 172 contained the 5p telomere (Event type: 00=4/16, 0=56/360, 1=9/49, 1o2T=0/49, 2T=57/286, 2A=0/11, 2G=44/75, 2C=1/7 and 2M=1/4). Proportion of 5p telomere fusion events were indicated above each bar.

Complex inter-chromosomal telomere fusion events (2C) were also detected (**Figure 4.15**). These were composed of distinct loci, involving more than one of the different types of telomere fusions that may have resulted from cycles of BFB.

4.4.3 Variety of TVR identified from the intra analysis

Telomeres are composed of TTAGGG repeats; however, the telomere-proximal regions contain Telomere Variant Repeats (TVRs). TVRs are degenerated repeats such as TTGGGG, TCAGGG and TGAGGG (Allshire *et al.*, 1989, Baird, 1995), and other less frequent TVRs (Letsolo *et al.*, 2010, Lee *et al.*, 2014a).

In chapter 3, 5p TL was compared to 17p and XpYp using STELA (Baird *et al.*, 2003) and TCAGGG and TGAGGG variant repeat composition was characterised using TVR-PCR (Baird, 1995). In this chapter, a deep characterisation of TVR in patient CLL-B cells was possible by NGS of telomere fusion amplicons.

In order to evaluate telomere sequence content in patient CLL-B cells, all telomeric reads identified from the intra-chromosomal analysis (event type 00 and 0) were pooled together and the presence and proportion of distinct TVRs were counted (**Figure 4.16** and **Supplementary Table 13**). It was not possible to identify whether the telomeric sequence was from 5p, 17p or XpYp, therefore a comparison between the distinct telomeres was not achievable. Instead they were counted as total number of TVRs identified.

From 1950 telomeric repeats noted, 76% were the canonical TTAGGG and the remaining 24% consisted of 18 distinct TVRs. The most abundant variant repeats were TTGGGG, GTAGGG, TGAGGG, TCAGGG and TTAGGGG, constituting 18.41%. Followed by 3.85% consisting of TTCGGG, TTTAGGG, CTAGGG and the remaining 1.87% was made of TAAGGG, ATAGGG, TTAGGGGGG, TTAAGGG, TTTGGG, GTGGGG, TTAGTG, TTAGCG, TGGGGG, and CTGGGG sequences.

This study provides the first characterisation of TVR in patient's cells, indicating the variety and proportion of TVRs present in 9 patient CLL-B cells.

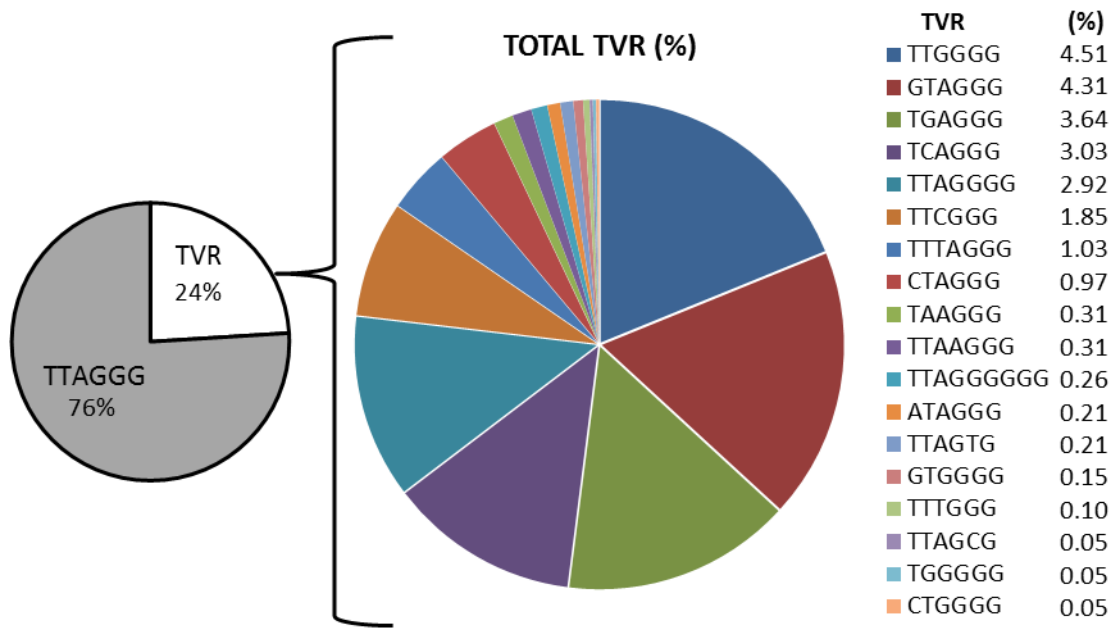


Figure 4.16. Telomere Variant Repeats identified.

Type and proportion (%) of TVRs identified from the characterisation of telomere fusion events. The list of TVRs was in order of abundance: from highest to lowest. Percentages recorded next to the TVR.

4.4.5 Subtelomeric deletion and asymmetry in sister chromatid and intra-chromosomal fusion events in CLL samples.

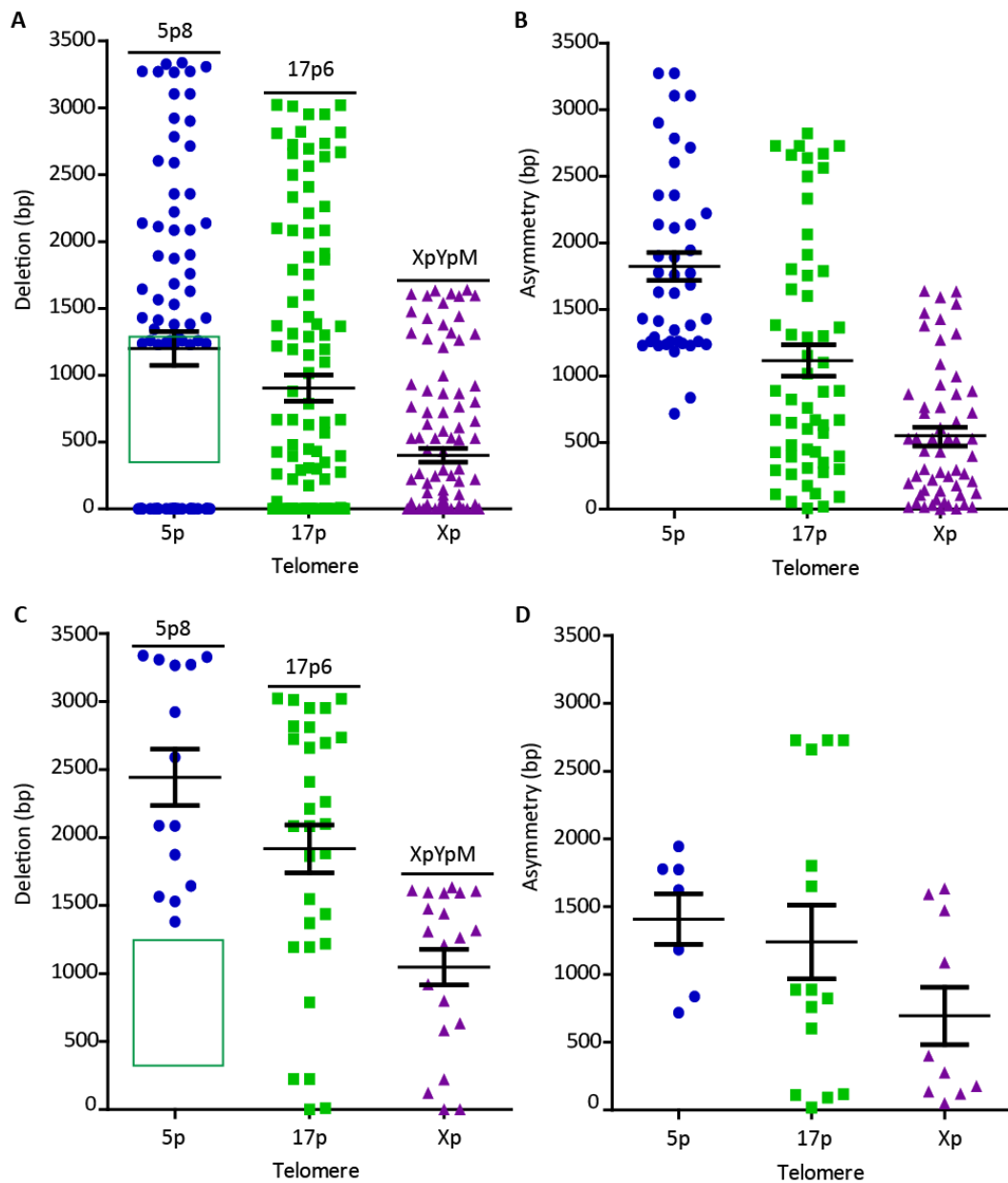
The characterisation of intra-chromosomal telomere fusions in MRC5 and HCT116 cell lines revealed the fusion of sister chromatids with different length, demonstrating an asymmetric deletion of each chromatid before ligation (Liddiard *et al.*, 2016). To assess whether this was true for patient CLL-B cells, subtelomeric deletion and asymmetry were investigated.

Deletion, which could either be caused by breakage or resection from the point of breakage, was investigated at intra-chromosomal fusion junctions by calculating the distance from the start of the telomere repeat sequences to the fusion junction for each of the chromatids involved in the fusion event. Subsequently, the difference in the degree of deletion between each sister chromatid was calculated to obtain the level of asymmetry. Sister chromatid deletion and the level of asymmetry for the 5p, 17p and XpYp telomeres (**Figure 4.17**) was calculated with events type 0 and 1 (subtelomere-telomere and sister chromatid fusion events) (**A-B**) and sister chromatid only (**C-D**).

Fusion junctions were evenly distributed between the fusion primer and the start of the telomere for the 17p and the XpYp subtelomeres. In contrast, a large interval in read alignments was found at the 5p subtelomere coincident with the location of the CpG island (**Figure 4.17A,C**).

It was not possible to determine whether telomere fusion events identified using the intra-chromosomal analysis were sister chromatid fusions or inter-chromosomal fusion events, therefore, asymmetry was more accurately determined using validated sister chromatid fusion events only.

Deletion at each sister chromatid extended up to the primer binding site with a mean of 2442bp, 1917bp and 1047bp for 5p, 17p and XpYp, respectively (**Figure 4.17C**). Asymmetry of sister chromatids (**Figure 4.17D**) was observed for 5p, 17p and XpYp with a mean of 1408bp, 1240bp and 695bp, respectively. The degree of asymmetry was significantly greater to the theoretical value 0 (one sample *t*-test, $p < 0.001$), demonstrating in patient CLL-B cells the fusion between sister chromatids of different lengths. No significant differences were found in the extent of asymmetry between the different ends (Kruskal Wallis, $p = 0.1661$).



4.4.6 Distinct microhomology utilisation at the fusion junction

To investigate the underlying DNA repair mechanisms (C-NHEJ or A-NHEJ) that may be associated with each type of event, microhomology at the fusion junction was documented (based on sequence of GRCh37 visualised on Ensembl).

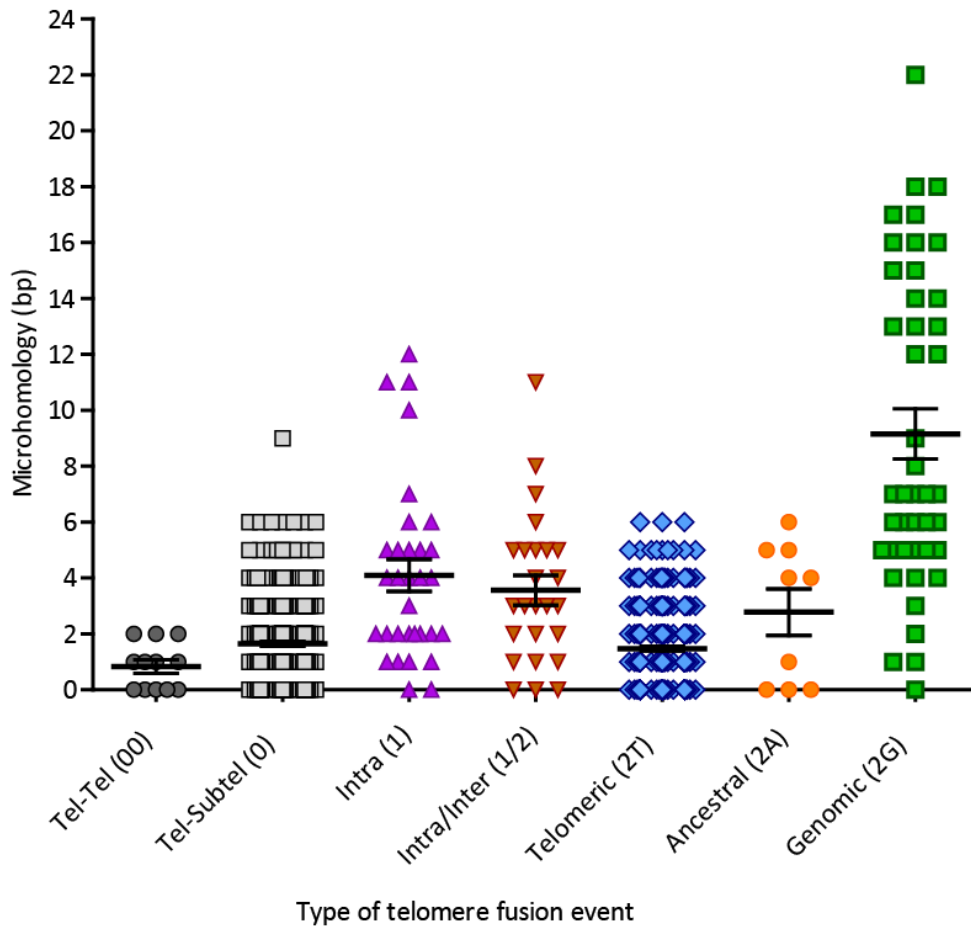


Figure 4.18. Distinct degree of microhomology at the fusion junction.

Microhomology (bp) at the fusion junction was compared for the distinct type of events: TTAGGG-CCCTAA (00), Subtelomere-TTAGGG (0), intra-chromosomal (1), intra- or inter-chromosomal of 16p-16p and 21q-21q families (1o2T), inter-chromosomal telomeric fusion events (2T), inter-chromosomal fusions with the ancestral telomere at 2q13 and inter-chromosomal fusions with genomic loci (2G). Inter-chromosomal fusions with mitochondrial DNA have been excluded as there were only 2 events with mapped fusion junction, one contained an insertion and the other MH=4.

Statistically significant differences in the extents of microhomology usage were observed between some of the different subgroups of telomere fusion events, shown in **Figure 4.18** and **Table 4.3**. Inter-chromosomal fusions with genomic loci (2G) (mean=9.12bp; n=43),

together with intra-chromosomal sister chromatid events (mean=4.09bp; n=32), showed the greatest extent of microhomology.

The microhomology usage identified for the subgroup intra/inter (1/2T) formed by fusion of the 16p-16p and 21q-21q families, was a combination of both intra- and inter-chromosomal telomeric fusions (mean=3.56bp; n=25). Followed by telomere fusions with the ancestral telomere at 2q13 (2A) (mean=2.78bp; n=9).

In contrast, very low or absence of microhomology at the fusion point was observed from inter-chromosomal telomeric fusions (mean=1.47bp; n=315), Tel-Tel (mean=0.83bp; n=12) and Tel-Subtel (mean=1.64bp; n=303) subgroups.

Long tracts of microhomology of up to 23bp, were observed at inter-chromosomal fusion junctions with genomic loci (**Figure 4.19**). When the usage of microhomology was >10bp, the sequence was enriched for the repeat unit of (AC)_n in 40% (6/15) of events that contained at least (AC)₅ (motif ACACACACAC).

Insertions were also observed at the fusion junctions in 6.28% (50/796): 23/50 for type 0, 4/50 type 1, 2/50 for type 1o2T, 19/50 for type 2T, 1/50 for 2A and 1/50 for 2M) but they were not identified at fusions with genomic loci. Insertions ranged from 1-21 nucleotides with a mean of 4.5 nucleotides and they were templated, untemplated and/or potential TVRs.

Table 4.3. Microhomology at the fusion junction

Type of event	00	0	1	1/2T	2T	2G	2A
Number of values	12	303	32	25	315	43	9
Mean	0.83	1.64	4.09	3.56	1.47	9.12	2.78
Std. Deviation	0.83	1.56	3.22	2.66	1.44	5.61	2.49
Std. Error	0.24	0.09	0.57	0.53	0.08	0.86	0.83
Type of event	00	0	1	1/2	2T	2G	2A
00: Tel-Tel	x	ns	**	**	ns	***	ns
0: Subtel-Tel	x	x	***	**	ns	***	ns
1: Intra	x	x	x	ns	***	ns	ns
1/2T: Intra/Inter	x	x	x	x	***	ns	ns
2T: Telomeric	x	x	x	x	x	***	*
2G: Genomic	x	x	x	x	x	x	ns
2A: Ancestral	x	x	x	x	x	x	x
Not all groups passed normality test. Non-parametric ANOVA Krustal Wallis $p < 0.001$ and Dunn's Multiple Comparison Test.							



Figure 4.19. Long tracts of microhomology at inter-chromosomal fusion junctions with genomic loci.

Examples of 5p inter-chromosomal fusions with genomic loci with different usage of microhomology (MH). **(A)** 5p-chr6p22.3, MH=18bp (DB60), **(B)** 5p-chr5p14.1 MH=18 (DB60), **(C)** 5p-chr2q14.1 MH=13bp (DB61), **(D)** 5p-chr3q13.32 MH=12 (DB66), **(E)** 5p-chr1q41 MH=9bp (DB65), **(F)** 5p-chr22q13.2 MH=6bp (DB59) and **(G)** 5p-chr19p13.12 MH=6bp (DB65). 5p subtelomeric sequence in capitals, genomic loci in green and microhomology at the fusion junction underlined.

4.4.7 Investigating telomere fusion activity with non-telomeric genomic loci

Telomere dysfunction is associated with increased genomic instability and disease progression in CLL (Lin *et al.*, 2010, Lin *et al.*, 2014); therefore a characterisation of inter-chromosomal telomere fusions with genomic loci was undertaken. This could reveal hotspots or provide information about the DNA sequence target of telomere fusion activity.

Inter-chromosomal fusions with genomic loci were identified across the 9 CLL patient samples. A total of 93 different locations were validated, of which 67.7% (63/93) had defined mapped fusion junctions (mFJ) and 32.3% (30/93) unmapped (uFJ). Some fusion junctions were not identified due to the reads not spanning this sequence (**Figure 4.20A**).

Distinct locations across the genome were incorporated into telomere fusions: 75 in different genomic loci (2G: 48mFJ, 27uFJ), 4 at mitochondrial DNA (2M: 2mFJ, 2uFJ), 11 at the ancestral telomere Chr2q13 (2V: 10mFJ, 1uFJ) and 3 genomic locations from complex events (2C: 3mFJ, 0uFJ).

Each location was depicted in a karyotype map shown in **Figure 4.20A**. However, the loci disrupted by telomere fusions were not randomly distributed as the number of mapped fusion junctions did not correlate with the chromosome length ($r^2=0.44$) or coding gene density of the respective chromosomes ($r^2=0.32$) (**Figure 4.20B**). However, loci with previously-reported copy number aberrations in CLL (Salaverria *et al.*, 2015) were found to be incorporated into telomere fusions including 2p15, 2p11.2 (2 events), 2q13 (11 events), 6q22.31, 11q22.2 and 18q21.32 (single events). In addition, a complex telomere fusion was detected involving 4 distinct loci including 13q14.2 that is frequently deleted in CLL (**Figure 4.15**).

Therefore, loci incorporated into telomere fusions were further investigated to reveal a potential association with the pathogenesis of CLL.

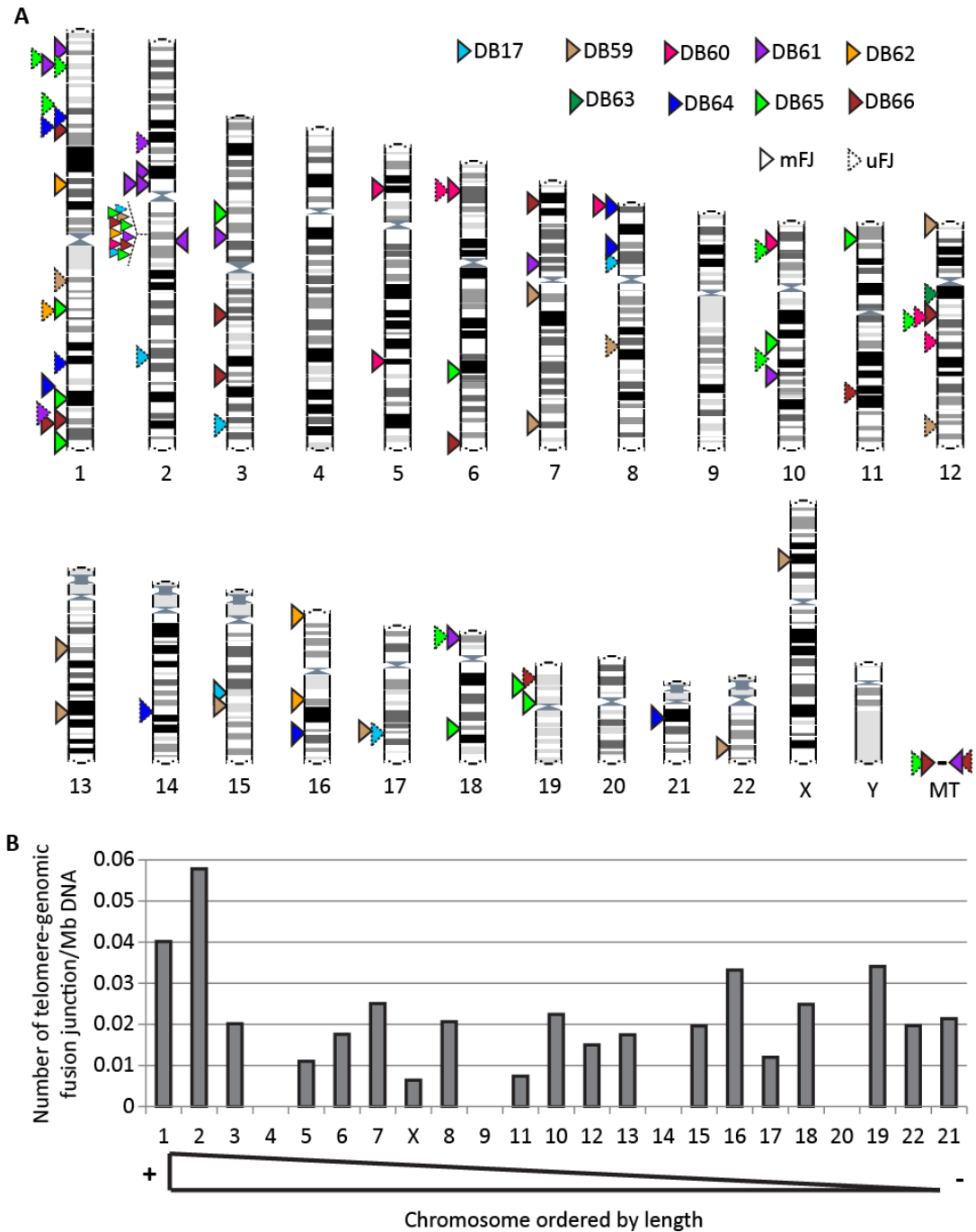


Figure 4.20. Inter-chromosomal telomere-genomic fusions from 9 CLL patient samples.

(A) Validated inter-chromosomal telomere fusion events on a karyotype map. Telomere-genomic, telomere-ancestral telomere 2q13 and telomere-Mitochondria DNA/Chr. All events were sequence-verified and BLAST-authenticated. Each colour represents a different patient sample. Continuous arrow-heads indicate mapped fusion junctions (mFJ), discontinuous arrow-heads represent unmapped fusion junctions (uFJ, location of the read represented). A total of 93 events comprising of 2G (n=75), 2C (n=3) (1 with two different locations), 2V (n=11) and 2M (n=4). Karyotype plot generated in Ensembl. **(B)** Number of validated inter-chromosomal telomere-genomic fusion junctions per Mb of DNA for each chromosome ordered by length (size obtained from Ensembl).

4.4.7.1 Telomeric fusions with mitochondrial DNA.

Four potential telomeric fusions with mitochondrial DNA (two with validated fusion junctions) were identified from the inter-chromosomal analysis (**Figure 4.21**). The list of genes disrupted was: MT-ND4, MT-ND5, MT-CO3, MT-TG, and MT-RNR1 (all exons).

- A** Tel_TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
 Telomere ←
 AAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT
 agtggaagaagaagagaggaagtaagtttaattatgccttttgggttgaggatgatggaggtg
 → MT:14120 (**MT-ND5**)
 gagattggctgctgaaattgttt_ChrMT
- MT-ND5:** Mitochondrially encoded NADH dehydrogenase 5, ENSG00000198786, exon 1/1, pc
- B** 16p_GTTTAAACTACATGCAGGAACAGCAAAGGAAATCCGGCAAATTTGCGCAGTCATTCTCAACACCGGC
 16:61426 ←
 CATGCAGCAAAA**TCAT**tactattctgcctagcaactcaactacgaacgcactcacagtgcataatcctctcaaggacttc
 → MT:11729 (**MT-ND4**)
 aaactctgctcccactaatagcttttgatgact_ChrMT
- MT-ND4:** Mitochondrially encoded NADH dehydrogenase 4, ENSG00000198886, exon 1/1, pc
- C** 16p_GCCCTGGGCACTATGGTTTGTAGCTGTACCCAGCGCTGCTTTGCTTGCTCTGTGACCCAGGCCAA
 16pfam; 16:64957-65056 ←
 GCTGCCTCACCTCTGGGCCAGTTCCCAT(...)(...)atactggcattttgtagatgtggttgctatttctgtatgtctccat
 → MT:9926-10025 (**MT-COA**, **MT-TG**)
 ctattgatgagggtcttactcttttctataaatgaccgttaactccaa_ChrMT
- MT-CO3:** Mitochondrially encoded cytochrome c oxidase III, ENSG00000198938, exon 1/1, pc
MT-TG: Mitochondrially encoded tRNA glycine, ENSG00000210164, exon 1/1, known Mt tRNA
- D** 5p_CACACACACTCACGTATATATATAAATCGTGGATAGCAATCTTATAGTAAGAAATTGTGACTTTTTA
 5:14701-14800 ←
 TAATCTTTGAAGAACTTTCCCATTTCTCACA(...)(...)taagctaaaactcacctgagttgaaaaactccagttgacacaa
 → MT:968-1067 (**MT-RNR1**)
 aatagacaacgaaagtggtttaacatatctgaacacacaatagctaagacca_ChrMT
- MT-RNR1:** Mitochondrially encoded 12S RNA, ENSG00000211459, exon 1/1, known Mt rRNA

Figure 4.21. Telomeric fusion with mitochondrial DNA.

Four examples of telomeric fusions with mtDNA. Fusion junction mapped for events **A** (DB61) and **B** (DB66), unmapped for **C** (DB65) and **D** (DB66). Mitochondrial DNA was represented in lowercase and pink, telomeric sequence was in capitals. Microhomology at the fusion junction was underlined and the insertion in highlighted in bold. Coordinate of the FJ and disrupted genes are indicated (pc: protein coding).

4.4.7.2 Fragility at the ancestral telomere at Chr2q13

A hotspot of telomere fusions with genomic loci was observed at Chr2q13 (**Figure 4.20, Figure 4.22**), consistent with the location of the ancestral telomere.

The inter-chromosomal telomere fusion analysis on 9 CLL patient samples revealed 11 distinct validated telomeric-2q13 fusions, 10 with mapped and 1 unmapped fusion junction. These events were detected in 7/9 CLL patient samples, with the 16p and 21q families and also with a TTAGGG-repeat sequence. The mean usage of microhomology at the fusion junctions was 2.78bp (as described in **section 4.4.4**).

Three of these fusion events involved the pseudogenes *DDX11L2* (DEAD/H (Asp-Glu-Ala-Asp/His) box helicase 11 like 2) and *WASH2P* (WAS protein family homolog 2), and the remaining were intergenic (**Figure 4.22A**).

4.4.7.3 Characteristics of genomic loci that fuse with dysfunctional telomeres

Validated telomere fusions with genomic loci (2G, n=75), in addition to those identified from complex events (2C, n=3) and fusions with mitochondrial DNA (2M, n=4) were selected for further characterisation.

The features of the genomic loci were analysed for all events (mapped/unmapped fusion junction, n=82) and separately for those with mapped fusion junction only for greater certainty (mFJ, n=54) (**Figure 4.23** and **Table 4.4**). Telomere fusions mainly disrupted introns in protein coding genes (43.90%, 51.85%), followed by intergenic regions (40.24%, 37.04%) and exons in protein coding genes (9.76%, 5.56%) in lower abundance. A small list consisted of pseudogenes (2.44%, 1.85%), long-intergenic non-protein coding RNA (LINC) (2.44%, 1.85%) and non-protein coding genes (1.22%, 1.85%), respectively.

Telomere fusions occurred within coding DNA more frequently than expected by chance. 57.4% of mapped fusion junctions were in introns and exons of protein coding genes, significantly higher (Chi-squared analysis $p=0.0024$) than the average 41.8% gene content in the human genome (based on the hg19 RefGene; Pruitt, 2014).

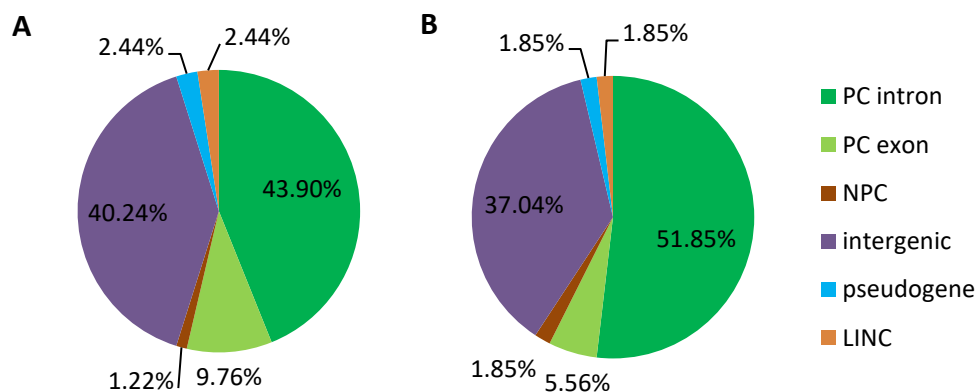


Figure 4.23. DNA features of genomic locus that fuses with dysfunctional telomeres.

Proportion of fusions with introns in protein coding (PC) genes (dark green), exons in protein coding genes (light green), non-protein coding (NPC) genes (brown), intergenic regions (purple), pseudogenes (light blue) or long intergenic non-protein coding RNA (LINC) (orange). **(A)** For 2G, 2M and 2C fusion events with genomic loci with mapped or unmapped fusion junction n=82. **(B)** Only for events with defined fusion junctions n=54.

Transcription regulatory elements such as promoters, enhancers and CTCF binding sites, were also found in some locations. Structural variations such as CNV were identified for some of the locations using information from the 1000 Genomes Project (Genomes Project *et al.*, 2015).

Telomere fusions have also been observed with repetitive DNA sequences. 18.3% (14.8% with mFJ) fused with Common Fragile Sites (CFSs), similar to the proportion of CFSs identified across the human genome (14.90%) (Fungtammasan *et al.*, 2012). In addition, 14.63% (9.25% with mFJ) of fusions were with Alu elements, similar to the abundance of Alu sequences estimated across the human genome (11%) (Lander *et al.*, 2001, Roy-Engel *et al.*, 2001) (**Table 4.4**). Also, 6 fusion junctions contained the (AC)_n motif (with n≥5).

Table 4.4. Description of genomic loci that fuses with dysfunctional telomeres.

FEATURE	Unmapped fusion junction	Mapped fusion junction
PC INTRON	<i>EIF2B5, C8A, ORAI1, DCAF6, RDH8, SLC39A12, ZNF678, TMEM63C</i>	<i>POLDIP3, TESPA1, CD8A, KIF26B, RORA, EVI5, FGGY, ZNF254, LPHN1, FAM78B, CSMD1, KIF13A, PCNXL2, NTF3, BEN domain containing 7, FTO, SLC30A10, C6orf123, VPS13D, ECE1, PTC3, HTR7, DGKB, NOX5, HDDC2, SHQ1, DMD, TBC1.</i>
PC EXON	<i>KIF13A, MIP, RALYL, MT-CO3, MT-RNR1, MT-TG</i>	<i>DDX18, MT-ND4, MT-ND5</i>
INTERGENIC	-	-
NPC	-	<i>CTC-575N7.1</i>
LINC	<i>LINC01090</i>	<i>LINC00441</i>
PSEUDOGENE	<i>RP11-353N4.6</i>	<i>RP11-520H11.1</i>
CFS	<i>FRA12E (12q24.31), FRA1F (1q21.2), FRAK1 (1q32.1), FRA2H (2q32.1), FRA3C (3q27.1), FRA1B x2 (1p32.2, 1p32.3),</i>	<i>FRA10D (10q22.1), FRA15A (15q22.2), FRA18B (18q21.32), FRA1D (1p22.1), FRA1B (1p32.1), FRA1I (1q44), FRA3D (3q25.2), FRA5E (5p14.1).</i>
ALU	<i>AluY x3 (2q32.1, 17q24.2, 8p12), AluSx3 (1q24.2), AluJr x2 (10p12.33, 19p13.2), FLAM_A (1p32.3)</i>	<i>AluY (15q22.2), AluSx3 x2 (19p12, 19p13.12), AluJr4 (1p22.1), AluJb (22q13.2).</i>

4.4.7.4 B lymphocyte specific-genes and other oncogenes are hotspots of telomere fusion activity in CLL

A total of 31 protein coding genes (n=28 introns, n=3 exons) (**Supplementary Table 15**) with mapped fusion junctions were identified (example **Figure 4.24**), these were loaded into the Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.8, (Huang da *et al.*, 2009a, Huang da *et al.*, 2009b)) and GSEA Gene Set Enrichment Analysis (GSEA, v5.2, (Mootha *et al.*, 2003, Subramanian *et al.*, 2005)) databases (<https://david.ncifcrf.gov/>; <http://software.broadinstitute.org/gsea/index.jsp>).

The DAVID Functional Annotation Tool was used to investigate enrichment in biological process gene ontologies (BP_GO) using Ensembl IDs; however, none of the categories were significantly enriched. Categories, gene count, *p*-value and Benjamini-Hochberg multiple testing corrected values are listed in **Supplementary Table 14**.

The Ensembl Gene IDs were converted to Entrez Gene IDs using DAVID Gene Accession Conversion Tool. The same list of 31 protein coding genes (**Supplementary Table 15**) was also submitted to GSEA and compared to all collections available in Molecular Signatures Database (MSigDB) (**Supplementary Table 16**) (Subramanian *et al.*, 2005) in order to find overlaps.

Enrichment in genes overexpressed in CD38⁺ patient CLL-B cells was found: *HTR7* (DB61), *KIF26B* and *LPHN1/ADGRL1* (DB65) (*p*-value 1.5e⁻⁶; FDR *q*-value 2.7e⁻²). These three genes were previously found to be upregulated in CD5⁺/CD19⁺/CD38⁺ CLL cells compared to CD5⁻/CD19⁻/CD38⁻ CLL cells derived from the same patient in a panel of 6 CLL patient samples (Pepper *et al.*, 2007). From the 9 CLL patient samples, DB64 and DB65 were CD38⁺ (>20%) and the remaining CD38⁻ (<20%).

Enrichment in genes with a promoter region containing binding motif that matches the annotation for the transcription factor 1 (TCF1) was found (*DMD*, *RORA*, *NTF3* and *HTR7* with *p*-value 2.51e⁻⁵ and FDR *q*-value 1.31e⁻²). Moreover, enrichment in genes overexpressed in breast and liver cancer was also identified. More information provided in **Supplementary Table 16**, **Supplementary Table 17** and **Supplementary Table 18**.

In addition, 19.35% (6/31) of genes disrupted by a telomere fusion were, to some extent, expressed in B lymphocytes (data obtained from GeneCards <http://www.genecards.org>): *CD8A* (CD8a molecule), *RORA* (RAR Related Orphan Receptor A), *TESPA1* (Thymocyte Expressed, Positive Selection Associated 1), *DMD* (Dystrophin), *NOX5* (NADPH Oxidase 5) and *NTF3* (Neurotrophin 3).

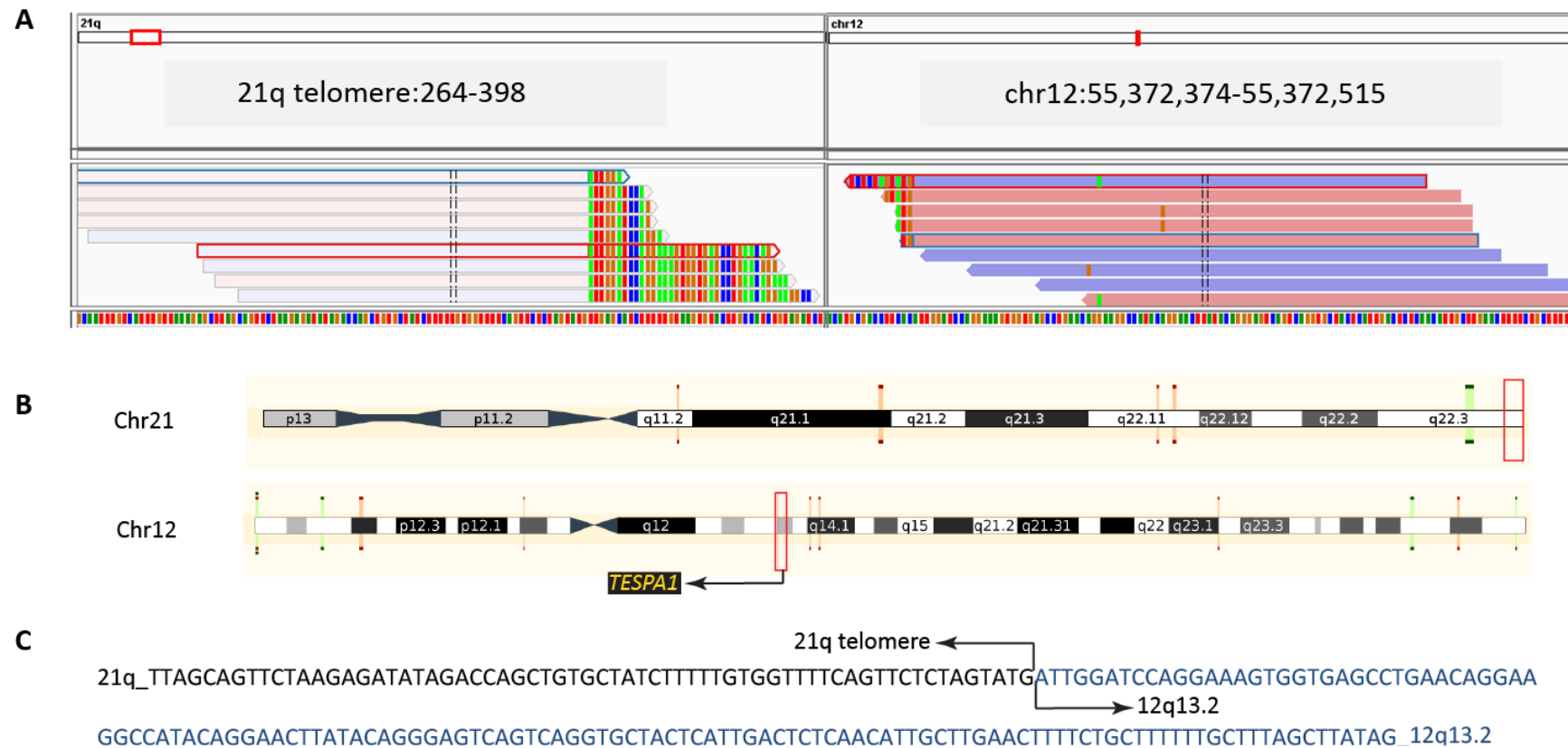


Figure 4.24. Telomere fusion with genomic loci 12q13.2.

Type of fusion 2G. **(A)** Paired-reads mapped for 21q telomere and 12q13.2 together with breakpoint visualised in IGV. Repetitive element and fragile sites indicated with the blue box. **(B)** Location of the reads for the telomere fusion event shown in Ensembl. Gene identified *TESPA1*: Thymocyte expressed, positive selection associated 1 (ENSG00000135426) as indicated by Ensembl/Havana **(C)** Sequence of the fusion event. The 21q telomeric sequence represented in black and the 12q13.2 in blue.

4.5 DISCUSSION

4.5.1 Higher frequency of bimodal telomere length distributions in CLL-B cells with the highest frequency of fusions

The proportion of samples with a bimodal TL distribution amongst the CLL samples with the highest frequency of fusions is surprisingly high when compared to the frequency of bimodality observed from the ARCTIC and ADMIRE cohorts: 66.67% (6/9) versus 3.98% (11/276), respectively.

A possible explanation could be the presence of multiclonality (two distinct subclones with different TL) in this subset of samples with more genomic instability and potentially a more aggressive disease. This hypothesis is supported by the evidence of intra-patient tumour heterogeneity in CLL. With the presence of multiple subclones that co-exist and can predict relapse particularly since chemotherapy is known to change the tumour composition and select for resistant subclones (Landau *et al.*, 2013, Landau *et al.*, 2014).

The bimodal distributions observed in these patients could also arise from biallelic TL variation; however, it is not possible to establish this without undertaking allele-specific STELA. Allele-specific STELA requires the presence of characterised sub-telomeric sequence polymorphism, this is known for XpYp (Baird, 1995); however, 45% of individuals are heterozygous and will thus not be informative in the two patients that exhibit bimodality at XpYp. A less likely explanation may simply be related with the purity of the sample (presence of other cell types) usually characterised by a very heterogeneous TL profile. To exclude this possibility, one of the patient samples was purified for CD19⁺ cells beforehand, revealing that the bimodal TL distribution was within the population of CLL-B cells. In addition, for most of the remaining bimodal samples, the absolute lymphocyte count (ALC) was similar or higher than the average from the 276 CLL patient samples from the ARCTIC and ADMIRE cohorts.

Taken together these results suggest an increased frequency of telomere dysfunction and fusion amongst a subgroup of CLL patients with intra-tumour heterogeneity. Telomere fusions might provide a mechanism for the acquisition of genomic rearrangements producing genetically heterogeneous tumour cells and contributing to multiclonal evolution.

Measuring TL using STELA provides valuable information about potential multiclonality within the sample that requires further investigation with other techniques given the impact of intra-patient tumour heterogeneity in disease progression and relapse. The co-existence of multiple clones could be assessed by checking the purity of CD19⁺ CLL cells in the sample, sorting CLL-B cells only, growing them in culture and measuring the TL profile at different time points. Multiclonality could be assessed based on the SNV from Whole Genome Sequencing (WGS) (further explored in Chapter 5).

4.5.2 Signature of intra- and inter-chromosomal telomere fusion events in a panel of 9 CLL patient samples

Results obtained from the paired-end NGS of telomere fusion amplicons in 9 CLL patient samples revealed a distinct signature of telomere fusion activity for each of the patients.

Complex telomere fusion events (n=7) incorporating up to 4 distinct remote loci were detected in patient CLL-B cells implicating wide-spread genomic instability. This was consistent with previous observations in MRC5 and HCT116 cell lines (Capper *et al.*, 2007, Letsolo *et al.*, 2010, Liddiard *et al.*, 2016). These results support telomere-driven BFB cycles in patient CLL-B cells, which can lead to distinct forms of genomic instability that are found in human cancers. Including gene amplification, gene loss by loss of heterozygosity (LOH) and non-reciprocal translocations, in addition to chromothripsis, kataegis, aneuploidy and whole genome duplication (Maciejowski and de Lange, 2017).

Dysfunctional telomeres require the acquisition of a new telomere to stop BFB cycles during a telomere-driven crisis. Some of the telomere fusion events identified from the 9 CLL patient samples were characterised by the presence of subtelomeric sequence followed by canonical TTAGGG-repeats in the same orientation. This could either be explained by the acquisition of a new telomere by translocation after BFB cycles (Murnane, 2006) or chromosome healing (Zschenker *et al.*, 2009). The latter event could have only been amplified with the fusion primers if the telomere became dysfunctional and fused again.

Reactivation of telomerase is required in 85% of human cancers for cellular immortalisation (Kim *et al.*, 1994). As the telomerase *hTERT* locus is proximal to the 5p chromosome end, chapter 3 focused on the adaptation of the telomere fusion assay to include the 5p telomere to investigate telomere dysfunction and fusion at this chromosome end. In this chapter, 172 telomere fusions involving the 5p telomere were validated, identifying sister

chromatid fusion events and inter-chromosomal fusions between 5p with distinct telomeres and several genomic loci. These results provide evidence of telomere dysfunction and fusion at the 5p chromosome end in CLL patients. Since translocation involving *hTERT* following telomerase upregulation have been observed in B-cell malignancies including CLL patients (Nagel *et al.*, 2010, Salaverria *et al.*, 2015), 5p sister chromatid fusion events may provide a mechanism for *hTERT* amplification.

In addition, in mantle cell lymphoma Chr5p15.33 is a recurrent breakpoint (Schilling *et al.*, 2013) that could be inappropriately repaired and incorporated into telomere fusions, leading to translocations of this gene near an enhancer or a transcription regulator. Further evidence comes from the escape from a telomere-driven crisis in MRC5^{HPCE6E7} cells after Chr5p amplification, including *hTERT*, and reactivation of telomerase (Jones *et al.*, 2014).

Further research is required to establish this association. A model system in cell lines could be designed to investigate activity at the 5p telomere that precedes 5p rearrangements and upregulation of *hTERT*, facilitating escape from senescence or crisis. A 5p TALEN/CRISPR technology could be used to induce DSBs at different locations of the 5p subtelomere and proximal to the *hTERT* locus. In addition, primers would be designed upstream each breakpoint to allow amplification of a fusion event near the breakpoint. This model would allow assessing whether instability at the 5p telomere facilitates the right rearrangement to amplify *hTERT* and escape crisis or senescence.

4.5.3 Variety of TVR identified from the intra-chromosomal analysis

The study of telomere sequence content in patient CLL-B cells revealed the presence of 18 potential TVR sequences: **TTGGGG**, **GTAGGG**, **TGAGGG**, **TCAGGG**, TTAGGGG, **TTCGGG**, TTTAGGG, **CTAGGG**, **TAAGGG**, **ATAGGG**, TTAGGGGGG, **TTAAGGG**, **TTTGGG**, GTGGGG, *TTAGTG*, *TTAGCG*, *TGGGGG*, and *CTGGGG*. TVRs highlighted in bold or underlined were previously reported in different cell lines by Lee *et al.*, 2014 and Letsolo *et al.*, 2010, respectively. Lee and colleagues suggested the presence of a different category made of sequences like: GG, GT, A, G, TT, GGG, TAGGGATAGAG, TTTGAGTAGGG, GCCAGGGTTAGAGT, TCAGAATTCAGAGGG, GTTATGG and TCAG. In addition, it remains possible that the least abundant hexameric repeats identified, highlighted in italics, were potentially a PCR or sequencing error arising from TTAGGG and TTGGGG, instead of a conserved TVR.

The study of TVRs is important. Although TVRs are located in the proximal regions (Allshire *et al.*, 1989, Baird, 1995), they can be found throughout the telomere (Lee *et al.*, 2014a). Their frequency and location is relevant because they decrease the binding affinity of members of the shelterin complex, having an impact on telomere protection and cellular fitness (De Lange, 2005b, Mendez-Bermudez *et al.*, 2009, Stohr *et al.*, 2010).

4.5.4 Asymmetric deletion at sister chromatids in intra-chromosomal fusion events

The uneven distribution of fusion junctions across the 5p subtelomere is consistent with the location of the CpG island and suggests that the GC-rich sequence may hamper the detection of 5p fusion events. In contrast, long-range fusion events were effectively captured across the 17p and XpYp telomeres.

The analysis of sister chromatid deletion caused by breakage or resection from the point of breakage in CLL-B cells, revealed asymmetrical processing of sister chromatids. This is consistent with results obtained from the NGS analysis of telomere fusions in colorectal cancer cell lines, postulated to arise from ligation of DNA strands that had been replicated differently (Liddiard *et al.*, 2016).

The ligation of sister chromatid arms of different lengths produced by stochastic DSBs at subtelomeric DNA and/or distinct length of resection, would result in asymmetric deletion.

4.5.5 Distinct degrees of microhomology at the fusion junction

Distinct degrees of microhomology usage at the fusion junction were identified for each type of telomere fusions event and may potentially be associated with specific DNA repair pathways.

In this study, intra-chromosomal and inter-chromosomal fusions with genomic loci presented the greatest usage of microhomology (4.09bp and 9.12bp, respectively), suggesting a role for LIG3/LIG1-dependent A-NHEJ. Also known as microhomology-mediated end joining (MMEJ), A-NHEJ is characterised by extensive resection to reveal 2-20bp (most common 4-6bp) of microhomology to stabilise the junction. In contrast, the absence or low usage of microhomology for telomeric inter-chromosomal fusion events (1.47bp) could be mediated by LIG4-dependent C-NHEJ, characterised by less processing of the DSB ahead of fusion. LIG4 does not rely on microhomology at the fusion junction but if required, the endonuclease Artemis is activated by the Ku70/80 and the DNA-dependent

protein kinase catalytic subunit (DNA-PKcs) complex, and trims slightly to reveal small sections of microhomology (<4bp) (Goodarzi *et al.*, 2006, Gu *et al.*, 2010, Chang *et al.*, 2015). However, some telomeric inter-chromosomal fusion events contained up to 6bp of microhomology, potentially suggesting a role for A-NHEJ. Increased microhomology usage has previously been observed at intra-chromosomal (3.01bp) compared with inter-chromosomal (1.62bp) telomere fusion events, mediated by the A- and C-NHEJ respectively Liddiard *et al.*, 2016). Furthermore, LIG3 A-NHEJ is a mutagenic DSB DNA repair mechanism that was essential to escape a telomere crisis (Jones *et al.*, 2014).

Homologous recombination (HR) requires high levels of homology and it has been implicated for some intra- and inter-chromosomal fusion events (Mao *et al.*, 2016). In contrast, Doksani and De Lange suggest a principal role of A-NHEJ at telomeres, even for DSBs within telomeric repeats (Doksani and de Lange, 2016). As microhomology usage is limited at telomeric repeats, in this study it may not be reliable to assess the DNA repair mechanism involved for events 00 (TTAGGG-CCCTAA, 0.83bp) and 0 (Subtelomere-TTAGGG, 1.64bp) based on microhomology.

For genomic DSB repair, A-NHEJ is the main contributor to chromosomal translocations (Iliakis *et al.*, 2015). Intriguingly, two distinct clusters of microhomology usage could be observed for inter-chromosomal fusions with genomic loci (0-9bp and 12-22bp). It remains unknown whether it may be indicative of the participation of specific repair mechanisms and the accumulation of fusion events through the years.

In addition to HR, another mechanism implicated in DSB repair that requires greater usage of homology than A-NHEJ is single-strand annealing (SSA) (>20bp). It is similar to A-NHEJ but it needs large resection to reveal extended homology and it does not use PARP and Pol θ . Instead, it uses RAD52 to mediate alignment and the ERCC1/XPF nuclease to excise the overhangs, causing large deletions before ligation (Bhargava *et al.*, 2016). Uninterrupted homologous sequences favour SSA repair, which can lead to chromosomal rearrangements. A potential role for SSA in a subset of malignancies has been suggested given the enrichment of repetitive elements in several cancer-related genes and in genomic rearrangements (Elliott *et al.*, 2005, Belancio *et al.*, 2010, Zhang *et al.*, 2011b, Bhargava *et al.*, 2016).

In addition, the presence of the repetitive unit (AC)_n in 40% of inter-chromosomal fusion junction with genomic loci with microhomology usage greater than 10, might support the implication of SSA.

Therefore, inter-chromosomal telomere fusions with genomic loci could have been mediated by both A-NHEJ and SSA.

Alternatively, other potential mechanisms that require microhomology and lead to the formation of complex chromosomal rearrangements identified in human disease are: homologous recombination break-induced replication (BIR) and microhomology-mediated BIR (MMBIR) or fork stalling and template switching (FoSTeS). These are characterised by high usage of microhomology and templated insertions with duplications or triplications (Malkova and Ira, 2013, Zhang *et al.*, 2009, Holland and Cleveland, 2012, Sakofsky *et al.*, 2015). However, insertions were not observed in telomere fusions with genomic loci.

Insertions at the fusion junction were identified in 6.28% of the total events with validated fusion junctions (50/796). Surprisingly, insertions were found in very low frequency for sister chromatid events (4/50). Instead they predominated in telomeric inter-chromosomal events (19/50) and in subtelomere with telomeric repeat fusion events (23/50) yet, the latter, could be part of rare TVRs instead of insertions. A-NHEJ DNA Pol θ can incorporate nucleotides including templated insertions (Wyatt *et al.*, 2016); however, this repair pathway is proposed to mediate intra- and not inter-chromosomal telomere fusions. Otherwise, those insertions could have been introduced by the C-NHEJ DNA polymerases Pol μ and Pol λ that can synthesise DNA in a template-dependent/independent manner (Ma *et al.*, 2004, Bebenek *et al.*, 2014). Alternatively, FoSTeS/MMBIR has also been implicated in sealing telomeres by introducing terminal duplications to cap telomere fusions which may explain the templated insertions observed in head-to-head fusions (Lowden *et al.*, 2011, Yatsenko *et al.*, 2012).

4.5.6 Investigating hotspots of telomere fusion activity with genomic loci

Telomere-genomic inter-chromosomal fusions were observed across the genome and were not associated with chromosomal size, suggesting a non-random distribution. Loci with previously-reported copy number aberrations in CLL were found to be incorporated into telomere fusions. Moreover, two hotspot locations were identified on ChrM and Chr2q13 locus.

4.5.6.1 Telomeric fusions with mitochondrial DNA

Telomere fusions detected with mitochondrial DNA may potentially disrupt genes required for energy availability. One possible mechanism is a telomeric fusion with genomic loci

where there has previously been a nuclear mitochondrial (mt) DNA (NUMT) insertion, process called numtogenesis, that can be historical in the human genome (germline) or specific to the patient sample (somatic). The number of germline NUMT insertions in the human genome is between 755-1105 (Mourier *et al.*, 2001, Dayama *et al.*, 2014). However, NUMT insertions are up to 4 times more abundant in cancer than in normal cells (Srinivasainagendra *et al.*, 2017). Therefore, it is more likely that telomere fusions incorporate somatic rather than historical NUMT insertions given the higher abundance of the first type in cancer cells.

Ju and colleagues have shown the presence of mtDNA fusions into gDNA in cancer cell lines and primary tumours, frequently combined with nuclear rearrangements. They propose a mechanism of somatic mtDNA nuclear transfer and the implication of replication-dependent DSB repair or NHEJ for nuclear integration (Ju *et al.*, 2015). Compared to normal B cells, CLL cells have increased mitochondrial mass; and higher mtDNA copy number associates with higher risk of CLL (Carew *et al.*, 2004, Jitschin *et al.*, 2014, Kim *et al.*, 2015). Therefore, another option may be a telomeric fusion with mtDNA that has migrated into the nucleus, or within micronuclei.

Dysfunctional mitochondria can promote telomere shortening and chromosomal instability via reactive oxygen species (ROS), by-product of ATP oxidative phosphorylation (Liu *et al.*, 2002). Moreover, telomere dysfunction activates p53 which represses the mitochondrial regulator peroxisome proliferator-activated receptor gamma co-activator (PGC)-network which promotes metabolic failure diminishing cellular fitness (Sahin *et al.*, 2011).

Unlike normal B cells, CLL cells present an altered lipid metabolism. Similarly to adipocytes and myocytes, they store lipids in vacuoles, produce energy from free fatty acids (FFA) and express genes related to the lipid metabolism. With increased proliferation rate, the metabolic program is adjusted. To meet the energy requirement, the number of mitochondria and ROS production is increased (Bilban *et al.*, 2006, Rozovski *et al.*, 2016, Jitschin *et al.*, 2014).

Telomere fusions with metabolic genes encoded in the mitochondrial genome included mitochondrially encoded 12S RNA (*MT-RNR1*), cytochrome c oxidase III (*MT-CO3*), tRNA glycine (*MT-TG*), NADH dehydrogenase 4 (*MT-ND4*) and NADH dehydrogenase 5 (*MT-ND5*). In addition to telomere fusions with nuclear NADPH oxidase 5 (*NOX5*) and the lipid metabolism, fat mass and obesity associated gene (*FTO*) were also identified, which have been implicated in cancer cell metabolism (Roy *et al.*, 2015, Liu *et al.*, 2017c).

In CLL, disease stage has been associated with different metabolic state (Koczula *et al.*, 2016). Taken together these results suggest that CLL cells with an increased proliferation rate, overexpress metabolism-associated genes to meet the energetic requirement, facilitating telomere fusions with such actively-transcribed loci and producing additional ROS that further damage telomeres.

In addition, it has recently been shown that aberrant mitochondria release cytochrome C that results in self-inflicted nuclear DNA DSBs calling upon the ATM/ATR DDR, which contributes to maintaining tumorigenesis (Liu *et al.*, 2017b). This may be a potential mechanism for breakage and recombination between mitochondrial DNA, nuclear genome and telomeres.

The implication of telomere dysfunction and fusion with mitochondrial DNA needs to be further studied as it may have an impact in malignant transformation in conjunction with age-related disorders.

As the mitochondrial genome is only 16.5Kb long, 4 primers could be designed across the ChrM in each orientation to amplify potential telomere-mitochondrial fusion events when incorporated to our telomere fusion assay. To amplify telomere-ChrM fusions only, nuclear DNA would be extracted to avoid PCR artefacts and primers would target ChrM-specific DNA not incorporated in historical/germinal NUMTs. Another method could be the combination of the study of telomere fusions using FISH and mtDNA insertions with mtFIBER FISH (Koo *et al.*, 2017).

4.5.6.2 Fragility at the ancestral telomere at Chr2q13

From the analysis of inter-chromosomal fusions, telomere fusions with the Chr2q13 locus were identified. This locus contains the site of an end-to-end fusion of two ancestral chromosomes of a common ancestor with the apes that form human chromosome 2 (Ijdo *et al.*, 1991) estimated to have occurred 0.74-4.5 million years ago (Stankiewicz, 2016).

Previous research from our group showed some rare telomere fusions with the ancestral telomere at 2q13-2q14 (Letsolo *et al.*, 2010). This study provides further evidence and suggests the interstitial telomere, previously identified as a common fragile site (Bosco and de Lange, 2012), is a hotspot for telomere fusions. However, it remains possible that telomere fusions with the 2q13-14 locus are overrepresented because of the proximity of the 16p and 21q (family of primers) binding sites on this locus.

Genomic deletions at Chr2q13 have mainly been associated with genetic disease. Deletion at this locus has been observed in several patients and associated with a variable clinical phenotype: facial dimorphisms, developmental delay, autism spectrum disorder and restrictive cardiomyopathy amongst other (Yu *et al.*, 2012, Hladilkova *et al.*, 2015, Yu *et al.*, 2016). Although there are no common rearrangements at this locus in cancer, some cases with a 2q13 deletion have been identified in colorectal and prostate cancer (Dong, 2001, Hoang *et al.*, 2013, Wang *et al.*, 2016).

Genome wide association studies (GWAS) have identified SNPs at 2q13, in the acyl-CoA oxidase like (*ACOXL*) involved in lipid metabolism and proximal to the B-cell lymphoma 2 (*Bcl-2*) anti-apoptotic family, associated with CLL risk (Berndt *et al.*, 2013, Berndt *et al.*, 2016). Genomic instability arising from telomere fusions with this locus could have an impact on CLL pathogenesis. To further investigate such events, primers specific to the 2q13-14 sequence only could be designed, and in addition to a 2q13-14-specific probe, introduced in the telomere fusion assay.

4.5.6.3 Characteristics of genomic loci that fuse with dysfunctional telomeres

Telomere fusions with genomic loci predominated in protein coding DNA, consistent with previous observations of telomere-genomic inter-chromosomal fusions in colorectal cancer cell lines (Liddiard *et al.*, 2016). A potential explanation may be that these regions are transcribed and therefore the open chromatin status, the DSBs to initiate transcription and the nearby presence of the DNA repair machinery (Kakarougkas *et al.*, 2014) facilitates telomere fusions with these loci. If the gene function is disrupted, most likely cell viability will be compromised and result in cell death. However, it remains possible that the disruption of tumour-suppressor genes might confer a selective advantage following clonal evolution.

Telomere fusions with transcription regulatory elements including enhancers, promoters and CTCF binding sites, could result in gene upregulation if juxtaposed. CTCF is particularly relevant since it has been implicated in TERRA transcription, telomere protection and DSB repair by HR (Deng *et al.*, 2012, Hilmi *et al.*, 2017) and it is suggested to contribute to leukaemogenesis by promoting cell survival and inhibiting apoptosis in ALL (Zhang *et al.*, 2014).

Telomere fusions were also observed with CFSs and repetitive elements. CFSs are known to initiate cycles of BFB (Coquelle *et al.*, 1997) and are sites of frequent amplification, deletion

and translocations in cancer (Arlt *et al.*, 2006, Burrow *et al.*, 2009). Alu elements are the only active SINE in the human genome and the most abundant transposable element representing almost 11% of the total genome (>1M copies) (Lander *et al.*, 2001, Chen and Yang, 2017). In addition, the abundance of the (AC)_n motif at the fusion junction could be partially explained because this motif is enriched in Class 1 initiation site (IS) that represents 44% of all IS. It is a low-efficiency and late replication origin (Cayrou *et al.*, 2015) that could cause more replication fork stall.

4.5.6.4 B lymphocyte specific-genes and other oncogenes are hotspots of telomere fusion activity in CLL

Consistent with the previous explanation, several genes disrupted by inter-chromosomal telomere fusions are known to be expressed in CLL-B cells. Surprisingly, three of the genes disrupted: *HTR7* (serotonin receptor) in DB61, *KIF26B* (kinesin family member 26B) and *LPHN1/ADGRL1* (latrophilin 1) in DB65, have been found overexpressed in CD38⁺ CLL cells (Pepper *et al.*, 2007). This is relevant to the pathogenesis of the disease since high levels of CD38 are associated with shorter overall survival (Durig *et al.*, 2002, Pepper *et al.*, 2007). In addition, *LPHN1* is an acute myeloid leukaemia (AML) biomarker (Maiga *et al.*, 2016) and the *HTR7* is a family member of *HTR1* which is a therapeutic target in AML (Etxabe *et al.*, 2017).

Furthermore, it is particularly interesting to highlight that 3 of the genes disrupted are directly involved in lymphocyte development and are implicated in CLL: *TESPA1*, *RORA* and *CD8A*. *TESPA1* (Thymocyte expressed, positive selection associated 1) is expressed in T and B lymphocytes and regulates the inositol 1,4,5-trisphosphate (IP₃R) calcium-dependent activation of signalling pathways playing an important role modulating immune function (Matsuzaki *et al.*, 2012). *RORA* (RAR-related orphan receptor alpha) is involved in lymphocyte development and inflammatory responses (Dzhagalov *et al.*, 2004) and has been found over expressed in CLL among other cancers (Baskar *et al.*, 2008, Daneshmanesh *et al.*, 2008, Daneshmanesh *et al.*, 2015). *RORA* is also a very large CFS gene (within FRA15A 15q22.2) susceptible to genomic instability and inactivated in many tumours (Zhu *et al.*, 2006). *CD8A* (Cluster of Differentiation 8a), the CD8 antigen is expressed in cytotoxic T lymphocytes but aberrant expression has been reported in low frequencies in patient CLL-B cells and carries an adverse prognostic impact in the disease (Kern *et al.*, 2012).

Other disrupted genes to highlight are *EVI5*, *DMD* and *POLDIP3*. *EVI5* (ecotropic viral integration site 5) is a regulator of cell cycle progression and cytokinesis. It has been

suggested to prevent exhaustion in pre-leukemic stem cells in Runx1-deficient mouse (Jacob *et al.*, 2010) and to cooperate with BCL6 (B cell lymphoma 6) transcription factor in B and T-cell lymphomas (Baron *et al.*, 2014). In addition, deletion of 1p22 comprising *EVI5* has been identified in over 20% of patients with multiple myeloma (MM) and low expression of this gene associates with worse prognosis in early stage patients (Hofman *et al.*, 2017). DMD (Duchenne muscular dystrophy), dystrophin is found to be expressed at low but stable level in B cells and upregulated in unmutated CLL cases which associated with shorter survival (Bilban *et al.*, 2006, Nikitin *et al.*, 2007). *POLDIP3* codes for the polymerase delta interacting protein 3 (also known as *SKAR*) that regulates cell growth by specific binding with S6 kinase 1, the ribosomal protein downstream the phosphatidylinositol 3-kinase (PI3K) and mammalian target of rapamycin (mTOR) signalling pathways (Richardson *et al.*, 2004). *NTF3* (neurotrophin 3) and *NOX5* (NADPH Oxidase 5) were also disrupted and have, to some extent, been associated with B lymphocytes. *NOX5* was expressed in B cells in Hairy Cell Leukaemia (HCL) but not in CLL (Kamiguti *et al.*, 2005). B cells are a source of neurotrophins that provide protective autoimmunity in the damaged nervous system; however, they express the neurotrophins NGF and BDNF but do not seem to express *NTF3* and *trkB* (Edling *et al.*, 2004). Other genes associated with cancer pathogenesis have been identified, including *HDDC2*, *NTF3*, *KIF26B*, *VPS13D*, *DDX18*, *DMD*, *KIF13A*, *EVI5*, *POLDIP3*, *TBC1D15*, *FOX5*, *FGGY* and *SLC30A10*.

It is possible therefore, that genes implicated in B cell development and CLL malignancy, that are actively transcribed, are hotspots for telomere fusions in patients with CLL. These results are consistent with Boulianne *et al.* who observed that B-cell lineage-specific genes implicated in B-cell development were hotspots for DNA-DSBs during progression to malignancy (Boulianne *et al.*, 2017). Thus the presence of telomere fusions with non-telomeric genomic loci could be partially explained by the fragility associated with highly transcribed regions and the susceptibility to breakage of replication fork stalls under replication stress at CFs, or the DSBs to initiate transcription and proximity of DNA repair machinery.

Telomere fusions with genomic loci could be detrimental to normal gene expression, DNA replication, cell cycle progression and ultimately to cell viability. However, they have the potential to recombine distal locations, disrupt tumour-suppressor genes or reactivate genes that facilitate immortalisation, including *hTERT*. Therefore, telomere fusions can confer a selective advantage and provide a mechanism for cellular transformation and malignant progression.

4.5.7 Summary

For the first time, the impact of telomere dysfunction and fusion in the CLL genome has been studied. Telomere fusion amplicons from 9 untreated CLL patients with the highest frequency of fusions were characterised using a specialised paired-end NGS approach.

Distinct types of telomere fusions were identified including intra- and inter-chromosomal telomere fusion events and inter-chromosomal fusions with genomic loci. The later consisted on telomere fusions with the ancestral telomere at Chr2q13 fragile site, mitochondrial DNA and genomic protein coding loci. Elevated cellular replication can result in augmented mitochondrial mass. In parallel to increased energy production, the leak of cytochrome C and other ROS provoking additional DSBs may explain the incorporation of mitochondrial DNA into telomere fusions. It is also possible that telomeres fused with NUMT insertions. In addition, telomere fusions with coding DNA, particularly CLL, B cell or cancer associated genes, suggest that loci that are actively transcribed are more prone to damage. Complex telomere fusion events involving multiple loci have also been identified, which could juxtapose distinct genes and provide a selective advantage to the cell.

Furthermore, asymmetric deletion at sister chromatids before ligation was observed, most likely arising from stochastic DSBs or differential resection of short dysfunctional telomeres. Telomere protection could also be compromised by the wide variety of TVRs detected, which decrease affinity of the shelterin complex. Moreover, distinct usage of microhomology has been reported and suggests the implication of specific DNA repair mechanisms. These include C-NHEJ, A-NHEJ and potentially SSA consistent with extensive resection to reveal microhomology. To sensitise cells to DNA damage, therapeutic combinations targeting these specific DNA repair mechanisms may be required, to sensitise CLL B-cell clones with ongoing telomere dysfunction and fusion.

Given the proximity of *hTERT* to the 5p telomere, the detection of 5p sister chromatid fusion events may provide insights into another mechanism of telomerase amplification. In addition, 17p intra-chromosomal fusion events may lead to the loss of *TP53*.

Overall, telomere fusions were characterised in untreated CLL patient samples. Telomere fusions drive genomic rearrangements that may have a deleterious effect and trigger cell death or provide a survival advantage and clonal expansion. These events shape the CLL genome and provide a source of tumour heterogeneity that modifies the course of the disease.

CHAPTER 5:

WHOLE GENOME SEQUENCING IN A PATIENT WITH CHRONIC LYMPHOCYTIC LEUKAEMIA

5.1 ABSTRACT

Intra-tumour heterogeneity and clonal evolution is a persistent problem impacting the course of cancer. High-throughput sequencing technologies have enabled the characterisation of cancer genomes from individual patients and this has revealed the extent of genetic complexity and heterogeneity. Ultimately this information will facilitate the selection of tailored treatments, potentially avoiding drug resistance and patient relapse (precision medicine).

In this chapter, tumour heterogeneity was investigated in a patient with a bimodal XpYp telomere length distribution and the highest frequency of telomere fusions from a cohort of 276 CLL patients with short telomeres. The bimodal distribution was consistent with the presence of two CLL B-cell clones with distinct telomere lengths, which was confirmed by the allele frequency distribution of SNP data obtained from Whole Genome Sequencing (WGS) of tumour and patient-matched control samples. The analysis of mutations revealed a signature consistent with non-canonical activation-induced cytidine deaminase (nc-AID) which could play a role in early development of the disease. Mutations in CLL driver genes (*TP53*, *POT1*, *CREBBP*, *BRAF* and *ATR*) and in the MAPK/ERK pathway were identified, in addition to novel mutations in *REV3L*, *POT1-AS1*, *ATR* that may have an impact in telomere dysfunction and fusion. Chromosomal rearrangements, including a 6q and 13q deletion arising from a 6q:13q translocation, were detected. Interestingly, 17p copy neutral loss of heterozygosity or uniparental disomy (17p CN-LOH or 17p-UPD) was identified. A homozygous *TP53* (p.His179Tyr) mutation associated with high-risk CLL was also present in the patient sample. It is proposed that excessive genetic instability may be deleterious for the cancer cells, which together with the equilibrium of the subclones with long and short telomeres, contribute to the patient's asymptomatic disease.

5.2 INTRODUCTION

CLL is a heterogeneous disease with a median age at diagnosis of 72 years. A significant proportion of cases remains asymptomatic and never requires treatment while others have an aggressive disease that may eventually become refractory to therapy. Two distinct subtypes are distinguished depending if the CLL-B cells have undergone somatic hypermutation of the immunoglobulin heavy chain (*IGHV*) gene. The first subtype, mutated CLL (mCLL), is associated with an indolent disease while the second subtype, unmutated CLL (uCLL), confers a worse prognosis (Hamblin *et al.*, 1999, Damle *et al.*, 1999). To identify distinct subtypes of CLL, the clinical CLL staging system (Rai *et al.*, 1975, Binet *et al.*, 1977) and the *IGHV* mutational status, are used. Moreover, complementary information has been obtained from various molecular markers including, amongst others, the levels of zeta associated protein (ZAP-70) and CD38 expression (Hus *et al.*, 2006, Durig *et al.*, 2002).

Over the years, a wide spectrum of cytogenetic markers have been identified and associated with a distinct impact on the disease, treatment options and survival. The majority of CLL patients (50%) have normal cytogenetics or isolated 13q14 deletion and, in general, are associated with a more indolent form of the disease (Dohner *et al.*, 2000). Trisomy 12 (a gain of the entire chromosome) is the second most frequent abnormality observed in 20% of CLL cases and confers a prognostic impact of intermediate-risk (Dohner *et al.*, 2000). In contrast, 11q22-23 deletion (disrupts *ATM* and *BIRC3*) is associated with high-risk and therapy with Fludarabine, Cyclophosphamide and Rituximab (FCR) is the standard treatment option in these cases (Tsimberidou *et al.*, 2009). Very high-risk is conferred by deletion at 17p13 (disrupts *TP53*) which correlates with a more aggressive disease. Response to conventional chemotherapeutic agents has generally not been successful for this subgroup. Other therapeutic strategies like B cell receptor (BCR) inhibitors (Ibrutinib or Idelalisib) or allogeneic hematopoietic stem cell transplantation need to be considered for use in these subgroups of patients (Tam and Stilgenbauer, 2015). A complex form of genomic instability associated with poor prognosis is chromothripsis that was first identified in CLL and results in a wide-range chromosome fragmentation and reorganisation in 2-3% of cases (Edelmann *et al.*, 2012, Puente *et al.*, 2015).

The extensive heterogeneity observed in CLL disease progression has been partially revealed by NGS studies. Significant common CLL driver genes have been identified including: *TP53*, *ATM*, *NOTCH1*, *MYD88*, *BIRC3*, *SF3B1*, *BRAF* and *POT1*. Interestingly, mutations affecting these genes confer a distinct prognosis and response to treatment

(Wang, 2011, Puente *et al.*, 2011, Quesada *et al.*, 2012). In addition, a significant long list of other potential candidates has been suggested to drive the disease (Landau *et al.*, 2015). *POT1*, encoding protection of telomeres 1, is involved in telomerase recruitment and telomere capping. Somatic mutations in this gene have previously been identified in CLL patients and are associated with telomere dysfunction (Ramsay *et al.*, 2013).

In addition to the inter-patient tumour heterogeneity, multiple CLL subclones may co-exist within a single patient (intra-patient), impacting the evolution and outcome of the disease. Cancer cells acquire genetic and epigenetic mutations and, as they are exposed to intrinsic (microenvironment) and extrinsic (treatment) pressures, the fittest clones are selected over time (Darwinian selection) (Nowell, 1976, Greaves and Maley, 2012). This clonal evolution can be linear with a single clone that acquires the mutations over time or branched with multiple coexisting subclones (Landau *et al.*, 2013, Ouillette *et al.*, 2013).

These subclones can compete for resources or exist in equilibrium. In patients known to have intra-tumour heterogeneity, the distinct subclones were in equilibrium in the absence of therapy. However, for patients that received chemotherapy a subclone dominated over time, changing the equilibrium to a competition (Landau *et al.*, 2013). Polyclonality is a common cause of relapse (Landau *et al.*, 2014) and mutations such as *TP53* have been suggested to give advantage to dominate after relapse (Ouillette *et al.*, 2013).

In conjunction with a different mutational profile, subclonal populations may exhibit a distinct telomere length (TL) distribution. In CLL, TL is an important prognostic marker where homogeneous, heterogeneous and biclonal TL distributions have been detected using STELA (Lin *et al.*, 2010). It needs to be assessed in independent individuals whether this heterogeneity is an indicator of distinct allelic TL distributions or multiple subclones.

5.3 AIMS OF THE PROJECT

The telomere length profiles are generally very homogeneous in CLL patient samples and bimodality is only observed in a minor subgroup (4% of 276 CLL patients) (Norris *et al.*, manuscript in preparation). However, Chapter 4 revealed that most CLL patients (67%) with the highest frequency of fusions presented bimodal TL distributions.

The purpose of this chapter was to investigate the presence of intra-patient tumour heterogeneity in an untreated CLL patient (DB17) with a bimodal TL distribution and the highest frequency of fusions. It is hypothesised that a subclone with short dysfunctional telomeres could be driving genomic instability and CLL in this patient. However, the patient's indolent disease might be explained from the equilibrium with a distinct subclone with long telomeres.

The aims of this chapter were as follows:

- To determine the presence of distinct clones in a CLL patient sample with a bimodal TL distribution.
- To investigate the distinct subclonal populations and characterise the CLL genome: single-nucleotide variations (SNV) including CLL driver mutations, the signature of mutations, kataegis, *IGHV* status, copy number variations (CNV), translocations and chromothripsis.
- To identify the cause for the increased frequency of telomere fusions and the patient's favourable outcome.
- To examine the incidence of POT1 mutations that disrupt the POT1/TPP1 interaction, and to explore telomere dysfunction in CLL patients with short telomeres.

5.4 RESULTS

5.4.1 Whole Genome Sequencing of a CLL patient sample with a high frequency of telomere fusions and bimodal TL distribution.

5.4.1.1 CLL patient sample with bimodal TL distribution

In Chapter 4, patients with bimodal TL profiles were over-represented in the cohort exhibiting the highest frequencies of telomere fusions. To assess whether bimodality was explained by a biallelic or a biclonal distribution, DB17 was selected for more detailed study. Telomere length for this patient was previously determined for the 5p, 17p and XpYp telomeres; in addition, STELA at the 2p, 12q and 18q telomeres was performed.

The DB17 TL distributions observed at XpYp, 2p, 12q and 18q were contrasted to those obtained from HT1080 clone 2. In addition to the bimodality observed at XpYp for DB17, very heterogeneous TL distributions potentially suggesting the presence of bimodal distributions were observed at 2p and 12q telomeres. These results were in contrast to the homogeneity observed at the distinct telomeres for HT1080 clone 2 (**Figure 5.1**).

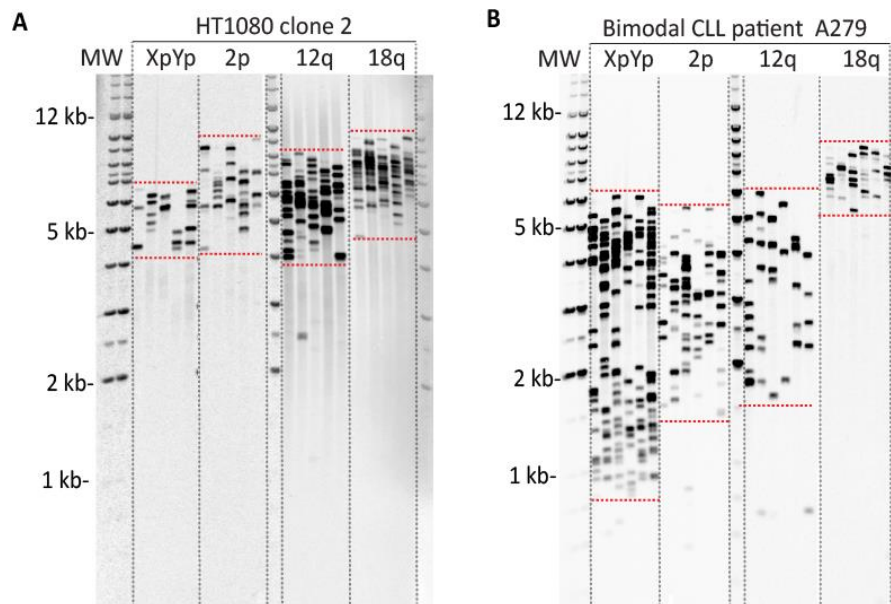


Figure 5.1. Telomere length at the XpYp, 2p, 12q and 18q.

STELA on different chromosome ends for **(A)** HT1080 clone 2 and **(B)** bimodal CLL patient DB17. Standard deviation for each TL profile indicated with red dotted lines.

TL at XpYp was the clearest indicator of bimodality for DB17 CLL-B cells. To interrogate the presence of biclonality, the evolution of the XpYp TL distribution was studied in DB17

samples obtained over 7 years processed similarly (**Figure 5.2**). XpYp STELA from DB17 patient sample at 2009 was provided by Dr Thet Lin. A stable TL can be observed for the longer TL distribution (A: XpYp TL>2.1Kb) across the 7 years, in contrast to the TL of the shorter distribution (B: XpYp TL≤2.1Kb) which erodes over time.

In addition, the proportion of telomeres detected within cluster B decreased at 2016 when compared to 2009, as can be observed from the A/B ratio (0.79 at 2009 to 1.07 at 2016). These results explain the increase in the mean TL at 2016 (2.56Kb) when compared to 2009 (1.99Kb) (**Figure 5.2**).

Taken together, these results suggest the presence of two clonal populations with distinct TL distributions (A and B), of which the clone A is stable through 7 years while the subclone B declines over time.

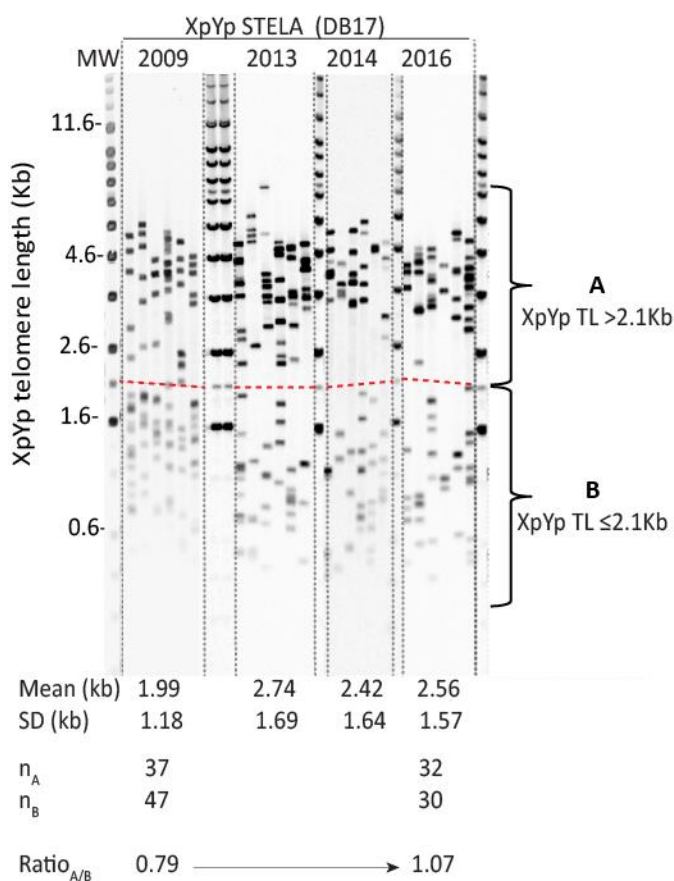


Figure 5.2. XpYp telomere length, 7 year follow up on a CLL patient.

Measurement of the TL using XpYp STELA on patient CLL-B cells throughout 7 years. Samples taken in 2009, 2013, 2014 and 2016. Hybridisation with the Telomere probe. Mean TL and standard deviation (SD) indicated underneath. Red dotted line indicates the separation of both TL distributions arbitrarily selected at XpYp TL 2.1Kb after subtracting distance to the primer 0.408Kb. Number of telomeres and ratio within clusters A and B for 2009 and 2016 indicated below. (2009 TL was provided by Dr Thet Lin.)

5.4.1.2 Preparation of the CLL patient sample for WGS

WGS was used to establish the clonal complexity of this patient's CLL and how this may relate to the TL distributions. Matched T cells (control) and CLL B-cell (tumour) samples were prepared for WGS.

Peripheral blood mononuclear cells (PBMC) were isolated from CLL patient DB17 fresh whole blood in heparin. CD19⁺ CLL-B cells were separated by CD19 microbeads using the autoMACS and after analysis with Accuri C6 flow cytometer, a purity of 97% CD19⁺ cells was observed. However, a pure CD3⁺ T cells fraction was not obtained using CD3⁺ beads. Therefore, another approach was performed.

CD19⁺ CLL-B cells and CD3⁺ T cells were purified from PBMC by flow sorting using the FACS-Aria III with CD19-APC and CD3-FITC antibodies (proportions before purification CD19⁺=61%; CD3⁺=9%). CD3⁺ T cells were selected as a control for their close lymphocytic origin to B cells and the feasibility to obtain the cell fraction from the same sample. Purity of the control fraction was checked after isolation (CD19⁺=1.2%; CD3⁺=97%) to ensure there was no contamination with cancer cells (**Figure 5.3**). A total of 2.5×10^6 T cells and 3×10^7 B cells were obtained for DNA extraction using Phenol/Chloroform.

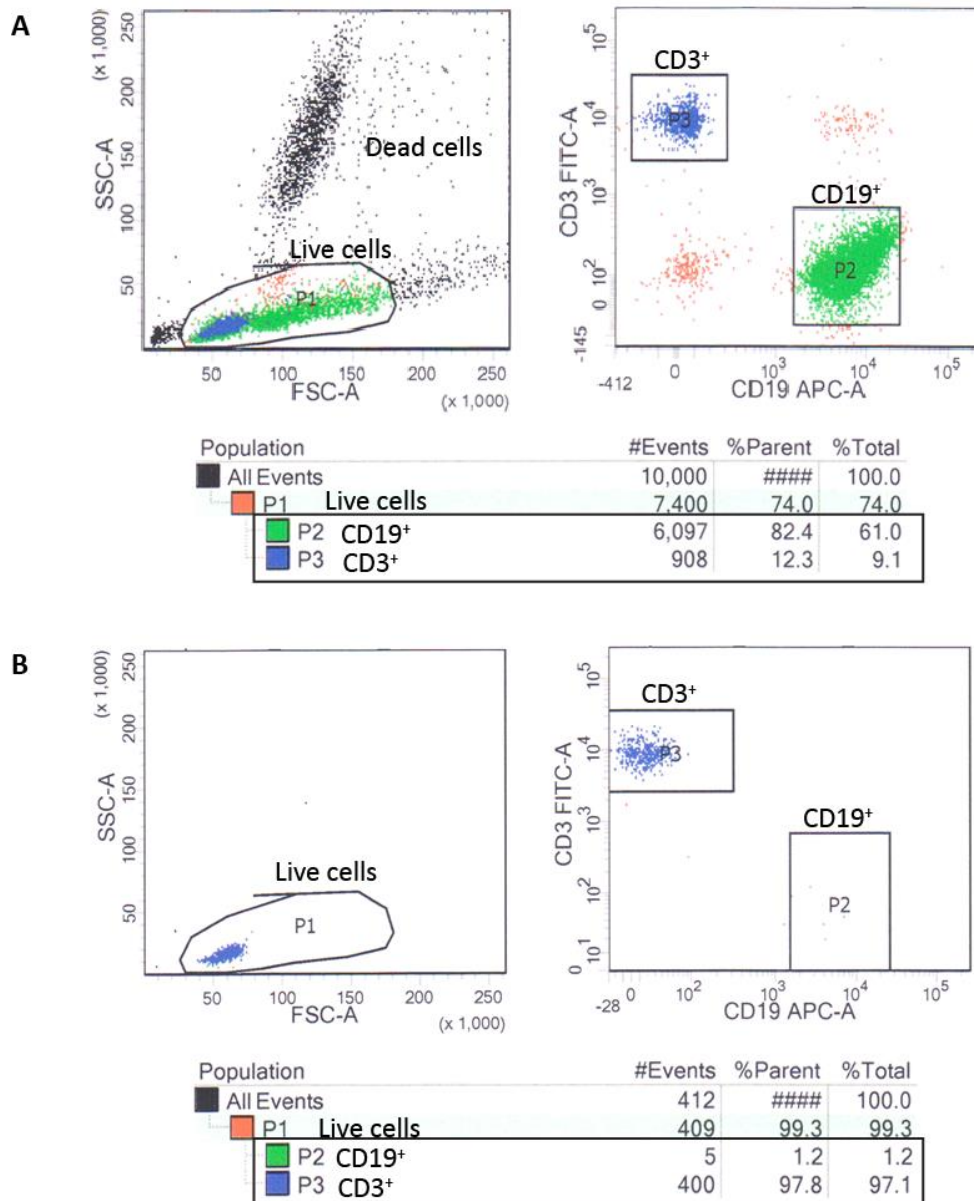


Figure 5.3. Isolation of CD19⁺ CLL-B cells (tumour) and CD3⁺ T cells (control).

(A) Purity of the patient sample before isolation of the CD19⁺ and CD3⁺ cells from PBMC by flow sorting using CD19-APC and CD3-FITC antibodies. (B) Purity of the control fraction, enriched in CD3⁺ lymphocytes. Proportion of CD19⁺ and CD3⁺ lymphocytes highlighted with black box.

To further verify the sample's purity, STELA at the 5p, 17p and XpYp was performed on the sorted CD19⁺ CLL-B and CD3⁺ T cells. The telomere length profile obtained from the control fraction was different from the tumour fraction and did not overlap, indicative of two distinct and pure populations. In addition, the XpYp TL profile did not correspond to any of the clusters observed for CD19⁺ cells, accordant with the presence of distinct subclones and not a distinct cell type (Figure 5.4).

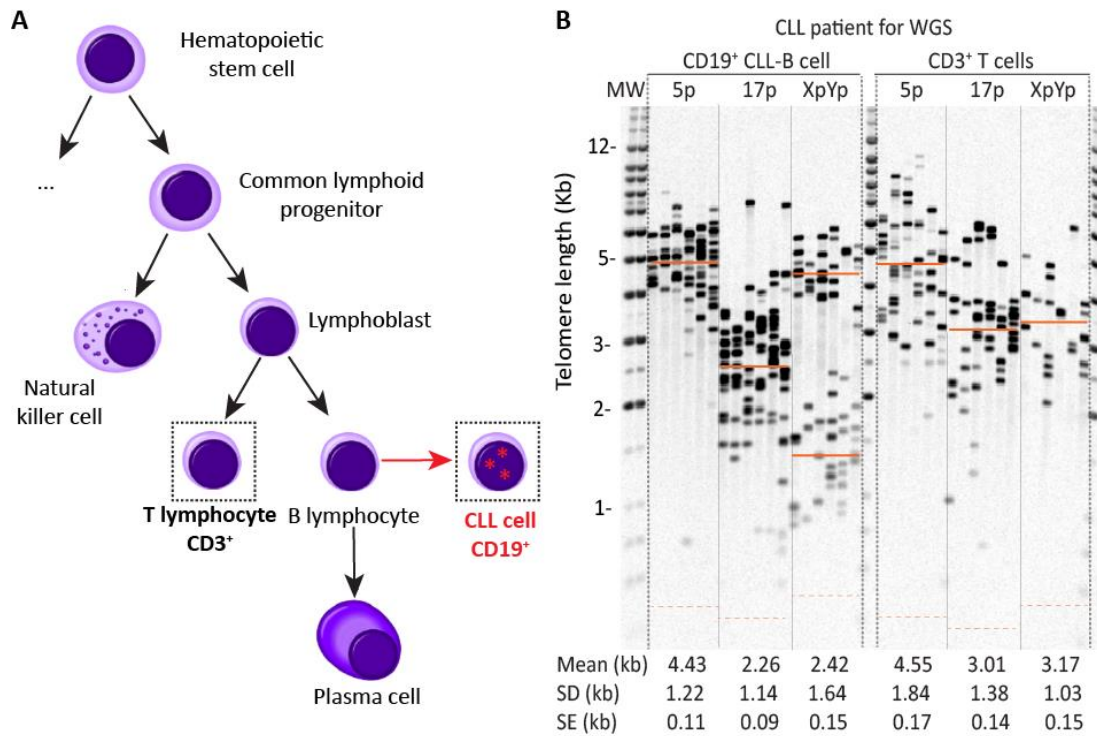


Figure 5.4. Comparison of the TL profile for the tumour and control fraction.

(A) Overview of a section of human haematopoiesis, highlighting the common lymphoid progenitor route. Common myeloid progenitor route (...) not shown. Within the dotted squared boxes, cells selected for tumour (CD19⁺ CLL-B lymphocytes) and control fractions (CD3⁺ T lymphocytes) for WGS.

(B) TL profile at 5p, 17p and XpYp telomeres for both tumour and control fractions.

WGS was undertaken at BGI Technologies with Illumina HiSeq2000, using 30µg of CD19⁺ cells gDNA for 60x coverage of the tumour genome and 2µg of CD3⁺ cells gDNA for 30x coverage of the control.

5.4.2 Intra-tumour heterogeneity detected in a CLL patient sample

5.4.2.1 Bioinformatics sequence analysis: Identifying SNVs

To identify SNVs unique to the tumour fraction a combination of two different methods to analyse tumour-control pairs were used: Mutect (Cibulskis *et al.*, 2013) and Somatic Sniper (Larson *et al.*, 2012) (**Supplementary Figure 3**). To increase accuracy and avoid false positives, the SNVs from the intersection of both callers were selected. To characterise somatic mutations and differentiate them from germline polymorphism, information from dbSNP database (<https://www.ncbi.nlm.nih.gov/SNP/>) and SnpEff (Cingolani *et al.*, 2012) was noted. These steps were performed by Dr Kez Cleal, the bioinformaticist in our group.

For a deeper exploration of the data and to use as a control, the analysis was also performed in reverse, using the CD3⁺ T cells as “R_tumour” and the CD19⁺ CLL-B cells as “R_control”.

	TUMOUR	CONTROL
STANDARD CONDITIONS	CD19 ⁺ CLL B cells (DB30)	CD3 ⁺ T cells (DB31)
REVERSED CONDITIONS	CD3 ⁺ T cells (DB31)	CD19 ⁺ CLL B cells (DB30)

The list of SNVs including SNPs and somatic mutations were documented in an excel file for further validation and analysis.

5.4.2.2 SNV analysis reveals multiclonality

To determine the presence of tumour multiclonality within the CLL-B cell fraction, the variant allele frequency (VAF) was assessed by calculating the fraction of sequence reads that supported each mutation. For a single clone with a heterozygous mutation, the VAF is 0.5 (1 mutation/2 alleles). In contrast, a biclonal tumour of which one clone carries a heterozygous mutation, the VAF is reduced to 0.25 (1 mutation/4 alleles). Consequently, a lower VAF is indicative of multiple clones (or genomes) within the sample.

The VAF for control CD3⁺ T cells resulted in a distribution peak at 0.5 associated with one population carrying polymorphisms normally present in the human population and not associated with disease, and not mutations. In contrast, the study of VAF for CD19⁺ CLL-B cells revealed two distribution peaks, one at 0.5 and the other at 0.25. These results indicated the presence of at least two distinct clones carrying clonal (present in all cells) and subclonal (present in the subclone only) mutations respectively (**Figure 5.5**).

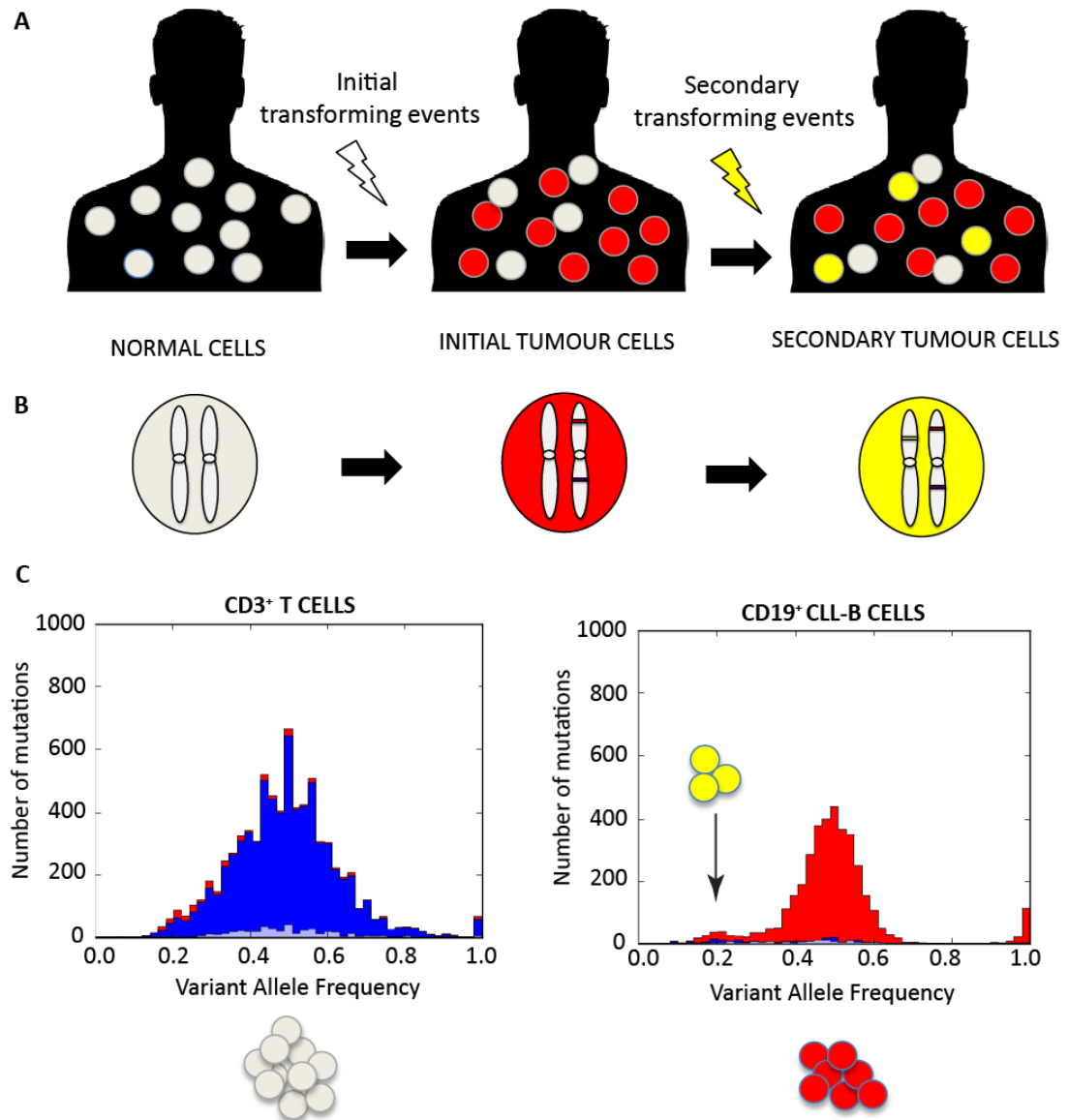


Figure 5.5. Intra-tumour heterogeneity detected from SNP analysis.

(A) Cartoon representation of evolution of tumour heterogeneity within a patient. Initial transforming event (white lightning bolt) gives place to initial tumour cells amongst normal cells (grey). Secondary transforming event (yellow lightning bolt) gives rise to secondary tumour cells (yellow). **(B)** Genetic mutations accumulated in tumour cells. Clonal mutations were those present in all cells originated from a primary event and subclonal mutations those that originated from a secondary event and are restricted to the subclone. **(C)** VAF obtained from the SNVs identified for CD3⁺ T cells and CD19⁺ CLL-B cells. Red: mutations, blue: polymorphisms.

5.4.2.3 Validation of tumour multiclonality by STELA and Sanger sequencing

For other studies DB17 CLL cells had previously been cultured *in vitro* over a 28 day period and DNA samples collected by Dr Ceri Jones. These samples were used to investigate if variation in the ratio of the two XpYp TL distributions over time in culture could support the presence of distinct clones. The STELA profile obtained at the XpYp telomere showed variation in the frequency of telomeres detected at both longer (A) and shorter (B) TL distributions between day 1 (baseline, BL) and the last day in culture (day 28). The number of telomeres detected in cluster A was maintained from BL to day 28. In contrast, the number of telomeres detected in cluster B at day 28 decreased to half compared to BL. This was represented with the ratio_{A/B} (0.45 at BL to 1.05 at day 28) (**Figure 5.6**). This variation in the number of telomeres detected was also reflected in the mean XpYp TL: 1.89Kb at baseline and 2.50Kb at day 28. The telomere dynamics observed in *in vitro* was consistent with those observed in DB17 over 7 years (**Figure 5.2**).

To verify whether this observation was related to a change in population frequencies indicative of multiclonal evolution, a subclonal heterozygous mutation in *IGF1R* chr15:99353338 G>A (26% of reads mutated) and a clonal heterozygous mutation in *POT1-AS1* chr7:124630153 T>G (52% of reads mutated) as a control, were amplified and re-sequenced at BL and day 28 (**Figure 5.7**).

The frequency of the *IGF1R* mutated allele diminished from 24% at BL to 12% at day 28, consistent with the loss of half the population with shorter telomeres at day 28 (ratio_{A/B}=1.05) from BL (ratio_{A/B}=0.45). In contrast, the frequency of the *POT1-AS1* mutated allele (control) was close to 50% and did not appreciably change over time in culture (BL: 53%, day28: 46%). The mutations were also Sanger sequenced for the CD19⁺ DB17 gDNA sent for WGS. Results from WGS sample did not differ from baseline (25% mutated *IGF1R*, 52% mutated *POT1-AS1*) (**Figure 5.7**).

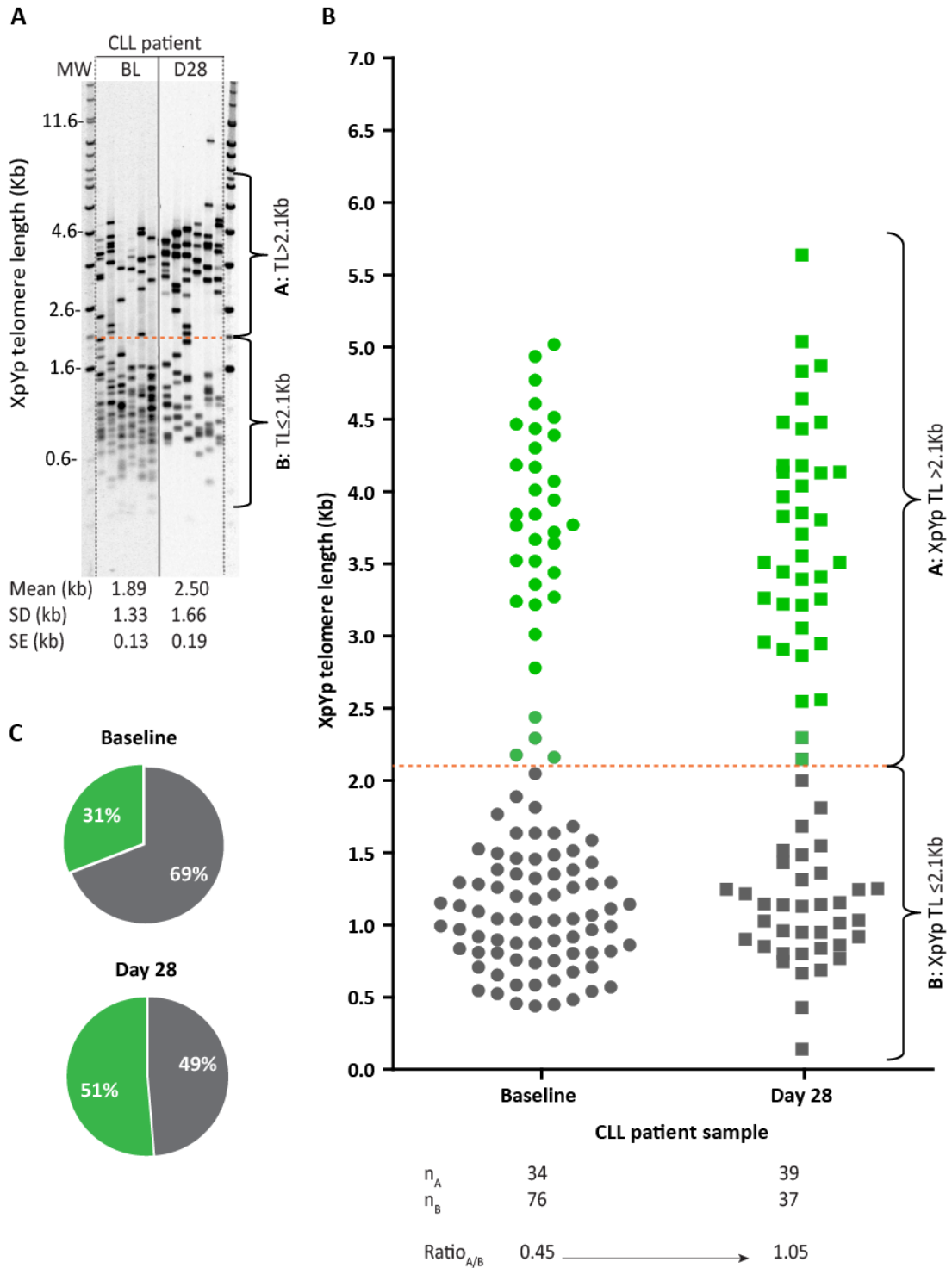
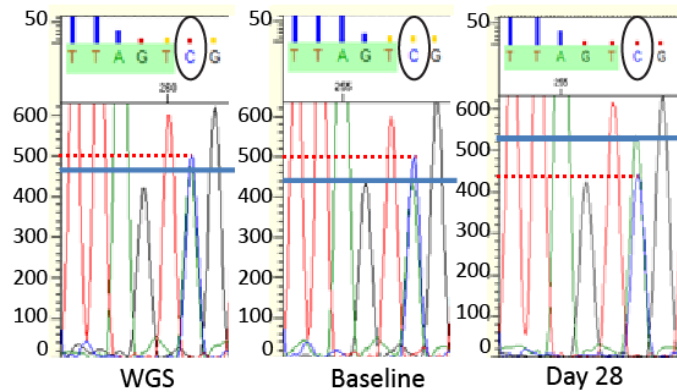


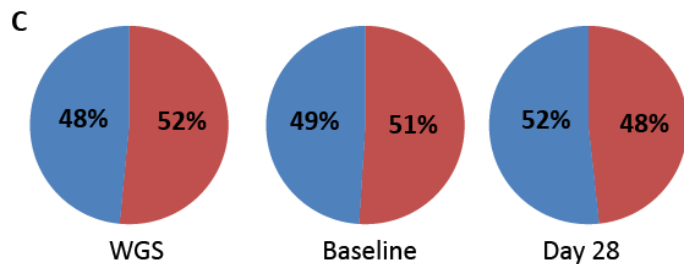
Figure 5.6. TL obtained from DB17 CLL cells in culture.

(A) XpYp TL at baseline (BL) and day 28 (D28) from cultured DB17 CLL cells. **(B)** Graphical representation. **(C)** Proportion of longer ($TL > 2.1Kb$) to shorter ($TL \leq 2.1Kb$) telomeres detected at baseline and day 28. Dotted line indicates arbitrary separation of both TL clusters: XpYp $TL \leq 2.1Kb$ in grey; XpYp $TL > 2.1Kb$ in green. Distance of the primer for XpYp STELA to the start of the telomere = 0.409Kb.

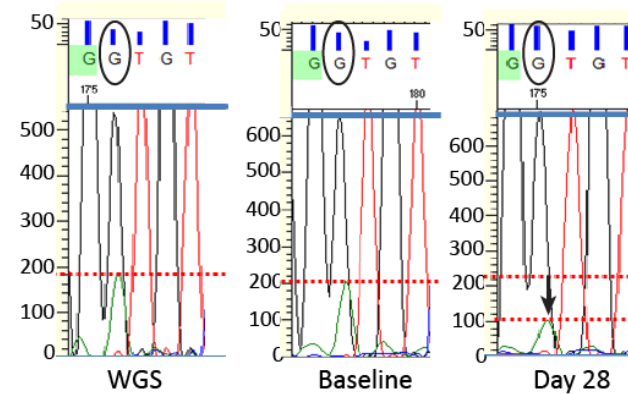
A CLONAL SNV: *POT1-AS1* (Chr7:124630153)



signal	WGS	BL	D28
intensity	500/470	500/480	495/530
Mutation	500	500	495
Reference	470	480	530
Total	970	980	1025
Mut (%)	51.55	51.02	48.25
Ref (%)	48.45	48.98	51.71
Total (%)	100	100	100



D SUBCLONAL SNV: *IGF1R* (Chr15:99353338)



signal	WGS	BL	D28
intensity	180/540	205/650	100/700
Mutation	180	205	100
Reference	540	650	700
Total	720	855	800
Mut (%)	25.00	23.98	12.50
Ref (%)	75.00	76.02	87.50
Total (%)	100	100	100

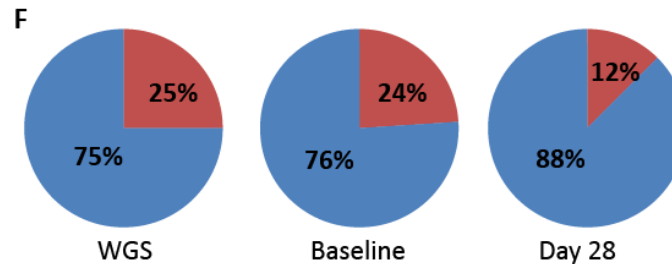


Figure 5.7. Validating multiclonality by sequencing mutations in DB17 at baseline and day 28.

(A-C) Sanger Sequencing of the clonal *POT1-AS1*: reference 48% T(A) and mutated allele: 52% G(C). (D-F) Sequencing subclonal *IGF1R*: reference 74% G(C) and mutated allele 26% A(T). For the WGS, baseline and day 28 DB17 samples. Position highlighted within a circle. (A, D) Electropherogram indicating intensity peaks. Maximum intensity peak indicated with blue line for the reference allele and red dotted line for mutation. (B, E) Table summarising intensity signal. (C, F) Proportion of reference and mutated allele for each sample.

5.4.3 Mutation in CLL drivers and genes associated with telomere dysfunction

5.4.3.1 Identification of mutations: Validation pipeline

From the list of SNVs identified within the tumour sample, 94% (3607/3821) were point mutations, with a frequency across the genome of 1.08 mutations per megabase of DNA ($3,607\text{bp} / 3,326,743,047\text{bp} \times 10^6 = 1.08 \text{ mutations/Mb}$).

In contrast, only 4% (309/7469) of SNVs identified within the control T-lymphocytes were mutations as 96% (7160/7469) were polymorphisms (**Figure 5.8**).

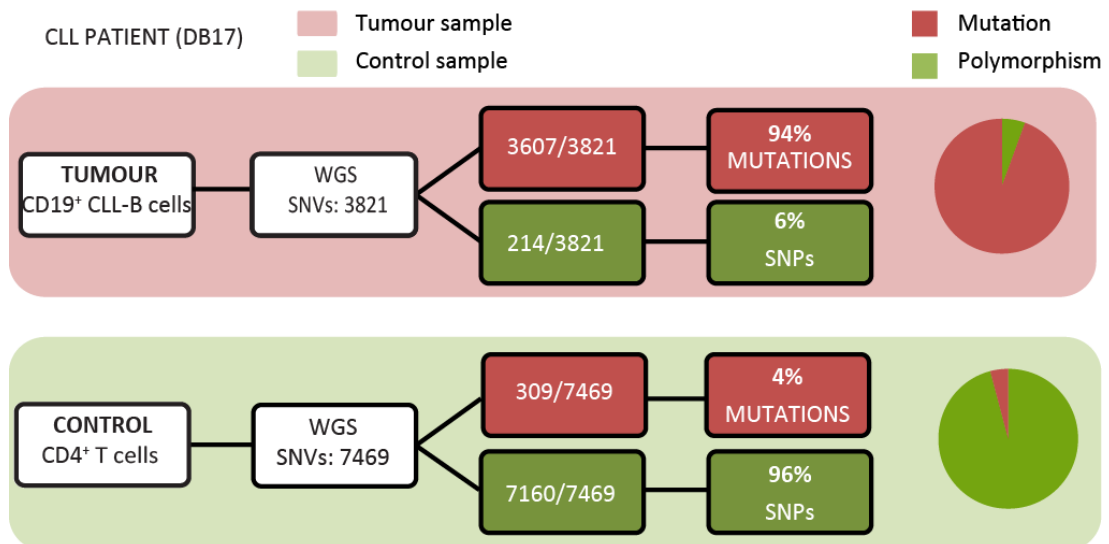


Figure 5.8. Summary of SNVs identified from the tumour and control sample.

Proportion of mutations and polymorphisms identified from the tumour and control fractions.

To confirm the accuracy of the mappers, selected SNVs were further validated by observation on IGV and BLAST-alignment of the reads using Ensembl (based on identity and E-value). The reference and mutated alleles, together with the mutation impact were noted. SNVs detected from the tumour fraction were further separated into clonal ($n=3655$) and subclonal ($n=165$) based on a VAF of >0.3 and ≤ 0.3 respectively, as previously observed in **Figure 5.5**. The certainty of the validation criteria (**Figure 5.9**) was confirmed by Sanger sequencing 8 mutations, 4 of which were subclonal (**Table 5.1**). The mutations were selected based on the location and predicted impact of the mutation on the gene, and the association of the gene with cancer or telomeres.

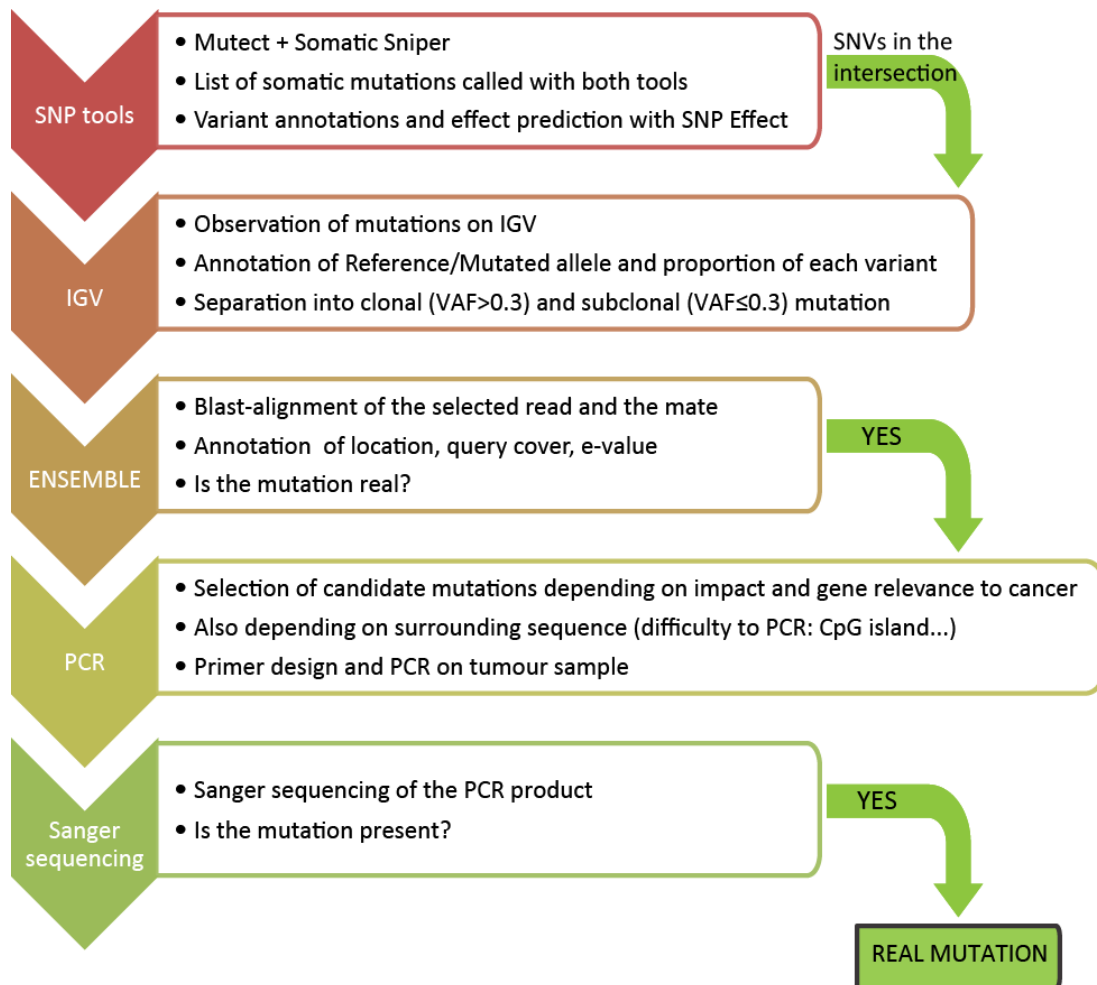
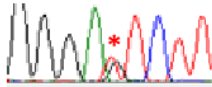
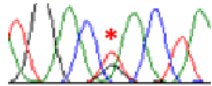
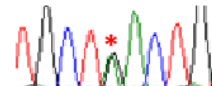
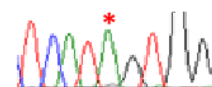
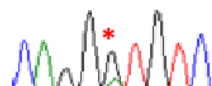
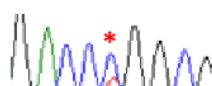
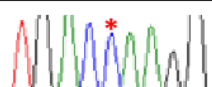
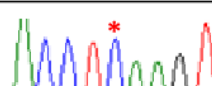


Figure 5.9. Pipeline for validation of SNVs.

Steps required for the identification, validation and sequence verification of SNVs detected from the tumour-control WGS sample. Identification using Mutect and Somatic Sniper, validation by observation in IGV and BLAST-alignment using Ensembl. Sequence verification with PCR and Sanger sequencing.

Table 5.1. Clonal and subclonal validated mutations.

	GENE NAME	MUTATED (%)	REFERENCE (%)	SANGER
CLONAL	<i>CREBBP</i> Chr7:140459419	G (61%)	T (39%)	
	<i>POT1-AS1</i> Chr7:124630153	G (52%)	T (48%)	
	<i>REV3L</i> Chr6:111695282	A (57%)	G (43%)	
	<i>TP53</i> Chr17:7578395	A (95%)	G (5%)	
SUBCLONAL	<i>IGF1R</i> Chr15:99353338	A (26%)	G (74%)	
	<i>EHMT1</i> Chr9:140589242	T (25%)	C (75%)	
	<i>TMPRSS15</i> Chr21:19845706	A (14%)	C (86%)	
	<i>BRAF</i> Chr7:140459419	T (16%)	C (84%)	

Clonal mutations are present in all cells while subclonal mutations are detected in distinct subclones. Proportion of reads with the mutated allele (red) and the reference allele (green) indicated. Snapshot of the Sanger sequencing trace and mutation shown ().

5.4.3.2 Impact of mutations: CLL drivers and the MAPK/ERK pathway

Somatic mutations were categorised based on the predicted impact of variants on genes: low (i.e. synonymous variant), modifier (i.e. intron variant), moderate (i.e. missense variant) and high (i.e. stop gained), or polymorphisms not associated with disease (predicted by SnpEff). A detailed description of the 3821 SNVs detected for the tumour CLL-B cells and 7469 SNVs for the control T-cells is summarised in **Supplementary Table 19**.

One high and 14 moderate impact clonal mutations predicted to effect a change at protein level were detected. 12 were located in protein coding DNA and 3 in regulatory sequence. A nonsense mutation (p.Gln1426*) in the protein coding gene *REV3L* changed glutamine (Gln) for a chain-terminating codon (*), predicted to create a truncated protein coding for 1426/3130 amino acids. Among the moderate mutations, 78.6% (11/14) triggered an amino acid substitution for a missense variant, affecting the following protein coding genes: *CD163*, *CREBBP*, *TP53*, *CD83*, *PCLO*, *FZD1*, *PLS3*, *AIFM1*, *PUS7L*, *MSR1* and *IGHV1-69*. The remaining 21.4% caused a sequence change at a glycosylation site in *SPG20*, *GHR* and *PKHD1*. The impact of the mutations on the biological function of the protein was assessed loading the Ensembl protein ID and the amino acid substitution onto the following bioinformatics online prediction tools: the Protein Variation Effect Analyzer PROVEAN (Choi *et al.*, 2012, Choi and Chan, 2015) and the Sorting Inolerant from Tolerant SIFT (Kumar *et al.*, 2009) from the Craig Venter Institute (**Supplementary Table 20**). In addition, a long list of 3080 modifier mutations and 508 with low impact were detected for all CLL-B cells, including a novel mutation at *POT1-AS1*.

As previously stated, mutations were divided on clonal ($VAF > 0.3$) and subclonal ($VAF \leq 0.3$), although it was impossible to assess which subclonal mutations were on the same population of cells. Neither high nor moderate impact mutations were subclonal. However, 137 modifier and 28 low impact subclonal mutations were identified. Given the gene association with telomeres, DNA damage response, CLL and oncogenesis, subclonal mutations highlighted included: downstream *ATR* and *MYB*, others affecting sequence features of *BRAF*, *MEIS1* and *RET*, and an intron variant on *IGF1R*.

To assess the main genetic differences present in all cells or characteristic for subclonal populations, functional analysis was perform independently with clonal ($VAF > 0.3$) and subclonal mutations ($VAF \leq 0.3$).

The Ensembl ID from 79 subclonal mutated genes (VAF \leq 0.3) were translated into Entrez gene ID using the DAVID conversion tool v6.8 (Huang da *et al.*, 2009a, Huang da *et al.*, 2009b) and submitted to the DAVID Functional Annotation Tool (**Supplementary Table 21**). Biological processes gene ontologies (GO_BP) enriched amongst this gene set included: signalling (38%), cell differentiation (26.6%), nervous system development (20%), positive regulation of biosynthetic process (15.2%), positive regulation of DNA-templated transcription (12.7%), cell migration (12.7%) and protein autophosphorylation (6.3%) (**Supplementary Table 22**). Similar results were observed from the list of clonal mutations. In addition, enrichment in the mitogen-activated protein kinase (MAPK) cascade was observed, including 52 distinct genes (5.5%) (**Supplementary Table 23**). These results may be relevant to malignant transformation since the MAPK pathway transmits signals to the nucleus that regulate cell growth, differentiation, migration and apoptosis amongst other.

The mutated genes identified within the CLL-B cells were compared to genes and pathways significantly enriched in CLL patients (Puente *et al.*, 2015, Guieze and Wu, 2015). Overlaps were identified for genes involved in the following key signalling pathways: DNA damage response, apoptosis, NF- κ B signalling, cell cycle, NOTCH1 signalling and B-cell signalling (**Figure 5.10**).

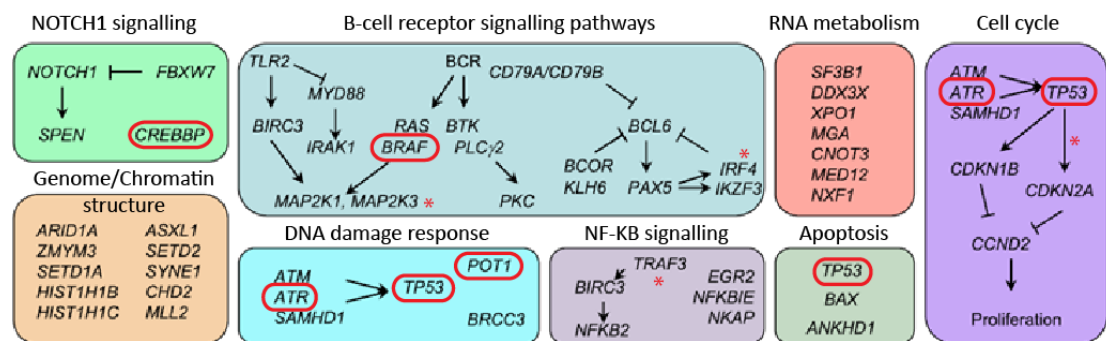


Figure 5.10. Significantly mutated genes and pathways in CLL.

Significantly mutated genes and pathways identified from Puente *et al.*, 2015 and Guieze *et al.*, 2015. Highlighted within red circle mutated genes identified from DB17 CLL-B cells (CREBBP, BRAF, ATR, TP53 and POT1). Red asterisk indicates family member or upstream/downstream of CLL significantly mutated genes, identified from the CLL patient sample. Cell cycle: PTPN12, PTPN2 (PTPN11) and upstream CDKN3 (CDKN1B, CDKN2A). NF- κ B signalling: TRAF2 (TRAF3) and downstream NKAP. B-cell signalling: upstream BRAFP1 (BRAF), MAP3K7, MAP3K13, MAP7, MAP1B (MAP2K1 and MAP2K3) and upstream IRF4. Genome/chromatin structure: ZMYM4 x2 (ZMYM3) and downstream HIST1H2APS6 (HIST1H1B). RNA metabolism: CNOT4 (CNOT3). Figure adapted from (Delgado *et al.*, 2016). Copyright (2016), with permission from Elsevier.

5.4.3.3 Telomeres and POT1 mutations

Protection of telomeres 1 (*POT1*) is a member of the shelterin complex that interacts with TPP1 and binds the single-stranded G-rich DNA overhang (**Figure 5.11A**) to cap telomeres and regulate telomere lengthening (Hwang *et al.*, 2012). *POT1* is essential for telomere stability. Mutations in this gene, most affecting the DNA binding domain, have been implicated in telomere dysfunction in 3.5% of patients with CLL (Ramsay *et al.*, 2013, Gu *et al.*, 2017) (**Figure 5.11B**). Recently, 7 *POT1* mutations predicted to disrupt the POT1/TPP1 complex have been reported (Rice *et al.*, 2017). To assess the impact of the interruption of the POT1/TPP1 complex, a cohort of 31 CLL patient samples with short telomeres (TL<3.81Kb), including DB17, were screened for the 7 mutations identified by Rice and colleagues (**Figure 5.11B, Supplementary Table 24**). These mutations were not identified in the cohort, suggesting that these events are not enriched in CLL patients with short telomeres.

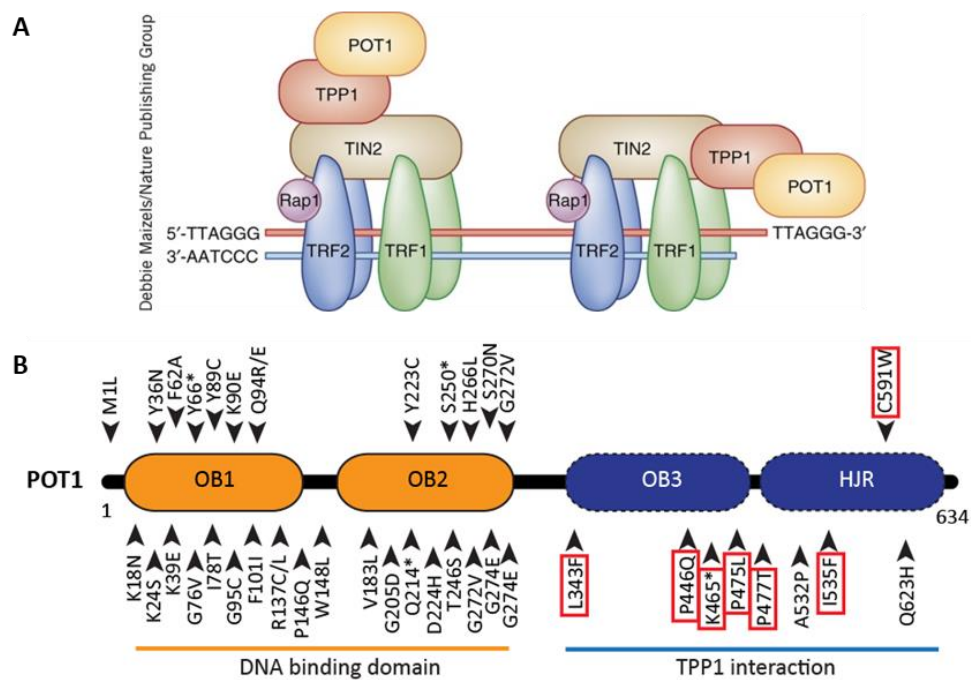


Figure 5.11. Point mutations identified in the human POT1 shelterin protein.

(A) Shelterin protein complex. Reprinted by permission from Macmillan Publishers Ltd: Nature Structural and Molecular Biology (Sarek *et al.*, 2015), copyright (2015). **(B)** Primary structure of the human POT1 protein. Somatic mutations identified in CLL or other cancer from different studies marked with black arrowhead. The DNA binding domains (OB1 and OB2) indicated in orange and the TPP1 interacting domain (OB3 and HJR) in blue. Highlighted within a red box, somatic mutations investigated for this study in 31 CLL patient samples with short telomeres (TL<3.81Kb). Figure adapted from (Ramsay *et al.*, 2013, Rice *et al.*, 2017, Gu *et al.*, 2017).

5.4.4 The signature of mutations in DB17 CLL-B cells is characterised by T>G transversions and associated with nc-AID

5.4.4.1 Transition/Transversion ratio

There are two types of DNA substitution mutations: base transitions (Ts) and base transversions (Tv). The first type interchanges the one-ring pyrimidines (C-T), or the two-ring purines (A-G). The second type exchanges one-ring pyrimidines with two-ring bases (A-C, A-T, G-C and G-T). There are twice as many possible transversions, but transitions are generated at a higher frequency as only an interchange of the same number of ring bases is required.

The whole-genome Ts/Tv ratio in human is reported to be between 2.1 and 2.3 (DePristo *et al.*, 2011) but the exact value is affected by GC content, proportion of CpG sites, allele frequency, natural selection, and other factors. From this data, the Ts/Tv ratio for the T cell control sample (reverse) was 2.15, similar to the Ts/Tv whole-genome expected ratio. In contrast, the ratio for the CLL B-cell tumour sample was 0.88, indicating that the CLL tumour fraction has a higher frequency of transversions than transitions (**Table 5.2**).

Table 5.2. Transition/Transversion ratio

SNVs	B-cell CLL		T-cell	
	count	%	count	%
A>G	505	13.22	1404	18.80
T>C	476	12.46	1353	18.11
G>T	433	11.33	308	4.12
G>A	415	10.86	1197	16.03
C>T	394	10.31	1141	15.28
C>A	355	9.29	318	4.26
T>G	264	6.91	285	3.82
A>T	258	6.75	265	3.55
T>A	253	6.62	260	3.48
A>C	249	6.52	292	3.91
C>G	118	3.09	339	4.54
G>C	101	2.64	307	4.11
	3821	100	7469	100

	B-cell CLL		T-cell	
Ts	1790	46.85%	5095	68.22%
Tv	2031	53.15%	2374	31.78%
Ts/Tv	0.8813		2.1462	

Table summarising the count and percentage of each type of substitution and the Ts/Tv ratio for the CLL-B lymphocytes and the control-matched T lymphocytes. Transitions indicated in orange and transversions in blue.

5.4.4.2 Signature of mutations

The pattern of somatic mutations can reveal information regarding the source or mutagen generating them (UV light, tobacco, ageing, etc.). Alexandrov and colleagues (2013) identified over 20 distinct signatures from 30 different cancer types. The method is based on 6 different base substitutions located in the middle of a tri-nucleotide motif providing 96 possible substitution mutations (Alexandrov *et al.*, 2013a, Alexandrov *et al.*, 2013b).

Although CLL has a low prevalence of somatic mutations (<1/Mb) compared to other cancer types (Melanoma ~15/Mb), three main signatures have been identified for CLL. From a cohort of 131 CLL samples, 86% presented signature 1B (age, C>T), 11% signature 2 (c-AID/APOBEC, C>T/G) and 17% signature 9 (nc-AID, T>G) (Alexandrov *et al.*, 2013a, Kasar *et al.*, 2015).

In agreement with the Ts/Tv ratio, the pattern of mutations identified for DB17 CLL-B cells was characterised by T>G transversions (**Figure 5.12**) and corresponded to Alexandrov's signature 9, one of the 3 signatures identified in CLL.

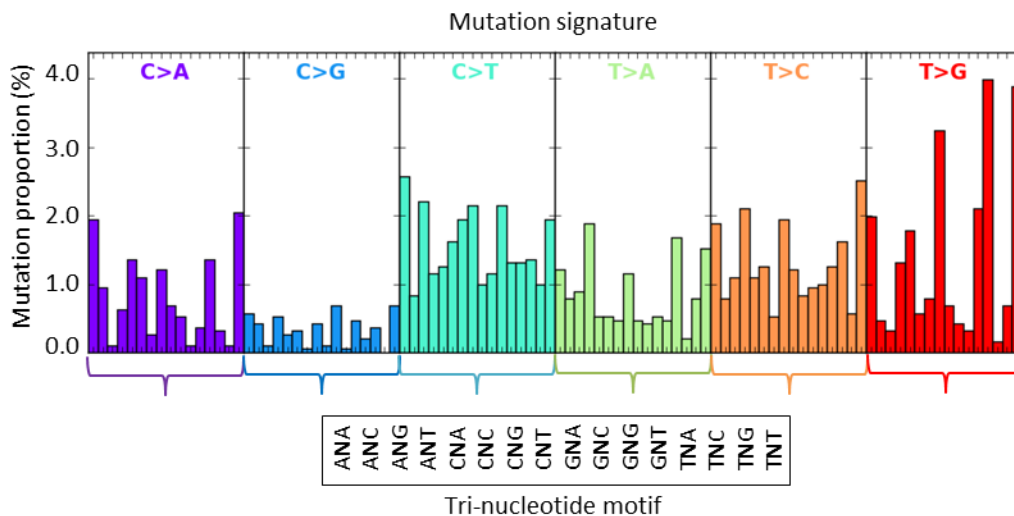


Figure 5.12. Signature of mutations.

Mutation signature found in CLL patient sample DB17 CLL-B cells. Pattern of mutations from left to right: repeats of Ax4, Cx4, Gx4, Tx4 for the first nucleotide, and ACGT for the third one as indicated in legend: tri-nucleotide motif. The middle base substitution (N) is indicated at the top.

5.4.4.3 Localised hypermutation is exclusive to the Immunoglobulin locus

In CLL, AID- and APOBEC-mediated hypermutation can lead to localised clusters of multiple mutations known as kataegis (Alexandrov *et al.*, 2013a, Rebhandl *et al.*, 2014). In contrast, nc-AID is characterised by non-clustered mutations (Kasar *et al.*, 2015). To further examine the signature of mutations, the presence or absence of kataegis was investigated.

This form of genomic instability is characterised by a group of 6 or more mutations separated by $\leq 1\text{Kb}$, therefore it was studied by mapping the genomic distance between each SNV (Figure 5.13).

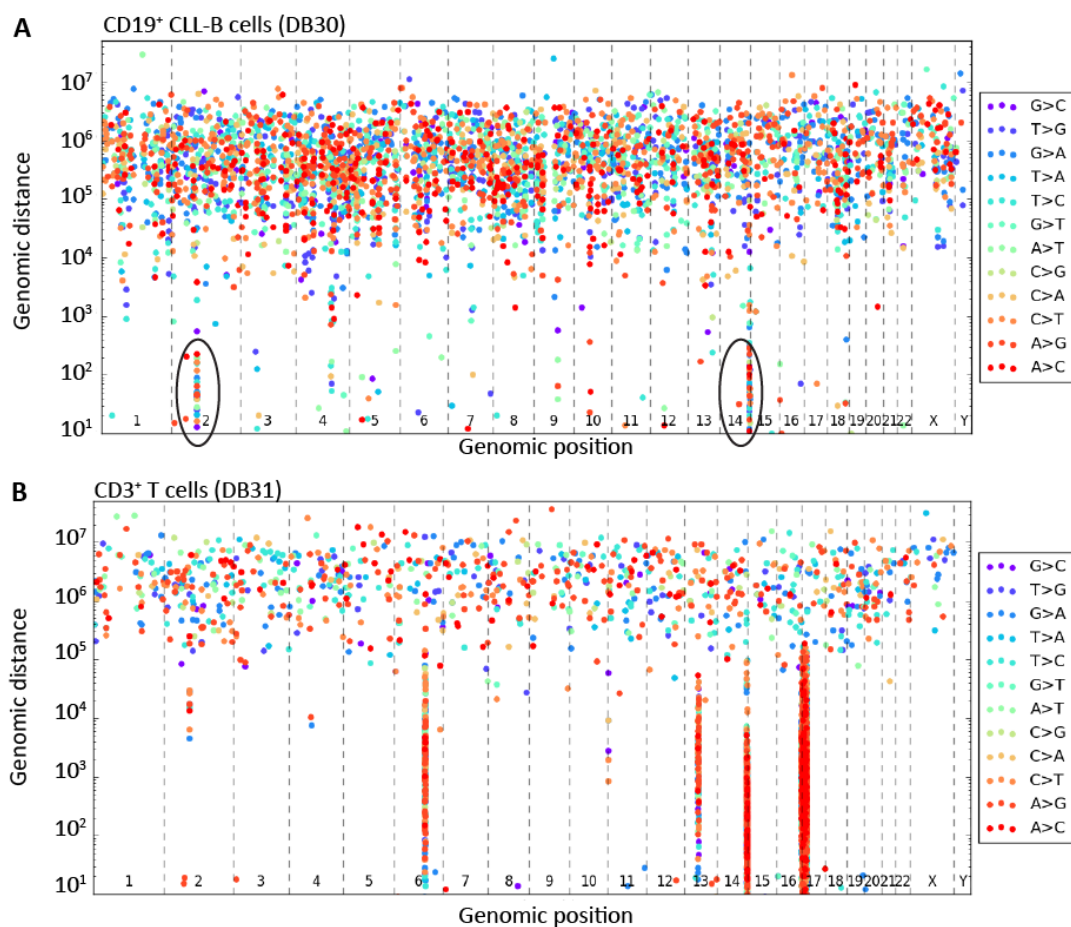


Figure 5.13. Proximity of mutation clusters (Kataegis).

Karyotype map plotting the genomic distance of SNVs to each other to identify clusters of mutations that may be indicative of kataegis for (A) DB30 and (B) DB31. Figure legend stating the change of base on the right. Cluster of mutations at Chr2p11.2 and Chr14q32.33 highlighted within a black circle. Intermutation distance indicated on the Y-axis (bp) and genomic position on the X-axis.

SNVs were evenly scattered across the genome for CLL-B cells (DB30) with exception of 2 potential clusters of mutations at 2p11.2 and 14q32.33 with intermutation distance below

1Kb. When compared to the karyotype map observed for the T cells (DB31), the 6q, 13q and 17p regions were identified. These co-localised with copy number deletions in the CLL-B cells at 6q and 13q, and CN-LOH at 17p described later in this chapter (**Figure 5.13**). In addition, a cluster of mutations at 14q32.33 was also identified for the T cells.

As the potential locus of kataegis located on chr2p11.2 was unique to CLL-B cells (DB30), this region was further investigated using IGV. 32 mutations were identified within 743Kb (chr2:89,159,650-89,160,393) in the tumour but not in the control sample (**Figure 5.14**). A coverage drop was observed at 3' from the cluster of mutations indicating a copy number loss. This was consistent with the 279Kb deletion detected using the Copy Number estimation by a Mixture Of PoissonS tool (cn.MOPS) (Klambauer *et al.*, 2012) at chr2:89160001-89440000 (**Supplementary Figure 4**). Over 30 reads supported a breakpoint at chr2:89,160,436 with the paired-reads mapping 282Kb downstream, providing confirmation of this deletion event.

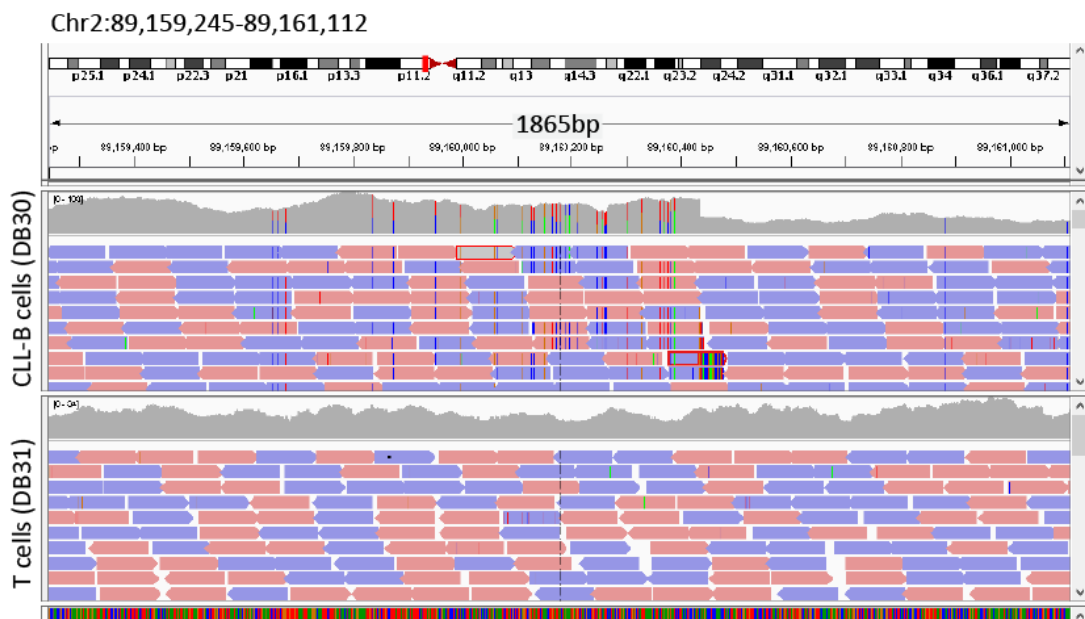


Figure 5.14. Kataegis and CNV at 2p11.2.

Cluster of 32 mutations identified within chr2:89,159,245-89,161,112 on CLL-B cells (DB30) followed by a CNV (deletion) as it can be observed from the drop in number of reads (coverage shown in grey) and breakpoint (read selected in red).

The immunoglobulin kappa joining 5 gene (*IGKJ5* exon1/1, ENSG00000211593) was located within the cluster of mutations at chr2p11.2. The CNV region included many immunoglobulin kappa variable and joining genes (*IGV* and *IGJ*) (Ensembl GRCh37).

At chr14q32.33, several clusters of mutations were observed within the tumour fraction along the genomic region that harbours Ig heavy locus genes (14:106032614-107288051).

In addition, deletions in that region were also identified in the tumour cells (**Supplementary Figure 4**), further supported with the cluster of SNVs present in the control fraction indicative of LOH (**Figure 5.13B**).

Altogether, these results were consistent with wide-spread mutations that are characteristic of nc-AID (signature 9). Mutation clusters were exclusive to the Immunoglobulin loci, distinctive of mature B lymphocytes that have undergone somatic hypermutation (SHM) (see **section 5.4.5**).

Despite the presence of specific copy number deletions, the absence of large-scale clustering of CNV across the genome suggested chromothripsis did not contribute to the CLL-B cell tumour genome (**Supplementary Figure 4**).

5.4.5 Mutated *IGHV* gene or multiple clones with diverse *IGHV* status?

Mature B lymphocytes contribute to adaptive immunity by producing immunoglobulins (Ig) that recognise foreign antigens. The process for generating Ig diversity is through somatic hypermutation (SHM) of Ig genes in the germinal centre.

Patients with unmutated *IGVH* and short telomeres have a more aggressive disease and shorter progression-free survival and overall survival (Strefford *et al.*, 2015). Assessing the *IGHV* status is a standard procedure in the clinic performed by RT-PCR and Sanger sequencing and compared to known germ-line genes.

The mutation status of DB17 was previously established using Sanger sequencing during the routine diagnostic work up. To assess whether multiple clones could be detected with diverse *IGHV* status, the *IGHV*₄₋₆₁ locus (Chr14:107,095,164-107,095,288) was visualised on IGV from WGS data (**Figure 5.15A**). The 9 SNPs previously detected by Sanger sequencing were confirmed consistent with DB17 presenting with mutated *IGHV* CLL (**Figure 5.15B**).

However, the VAF for each allele was ~0.35 mutated and ~0.65 reference, and mutated alleles were located in the same reads (**Figure 5.15A**, and **Table 5.3**). Recent *IGHV*-NGS analysis has revealed the presence of multiple subclones with different *IGHV* mutation status (Kriangkum *et al.*, 2015, Stamatopoulos *et al.*, 2017). There is the possibility that these results indicated the presence of multiple clones with differential *IGHV* gene status. However, it is likely that in agreement with the clinical diagnosis, the patient CLL-B cells present mutated *IGHV* and the observation in IGV represents allelic exclusion, which is the mono-allelic V(D)J gene recombination and SHM of the Ig cell surface antigen receptor (Vettermann and Schlissel, 2010, Fraenkel *et al.*, 2007).

Table 5.3. VAF of 9 mutations on the *IGHV*₄₋₆₁ locus for CLL patient DB17.

VAF	1	2	3	4	5	6	7	8	9
A (%)	2	0	0	<u>36</u>	0	<u>34</u>	0	0	<u>36</u>
C (%)	61	<u>41</u>	<u>36</u>	0	<u>33</u>	0	<u>35</u>	<u>35</u>	0
G (%)	1	0	0	0	0	66	0	0	0
T (%)	<u>35</u>	59	64	64	67	0	65	65	64

*Mutated allele underlined.

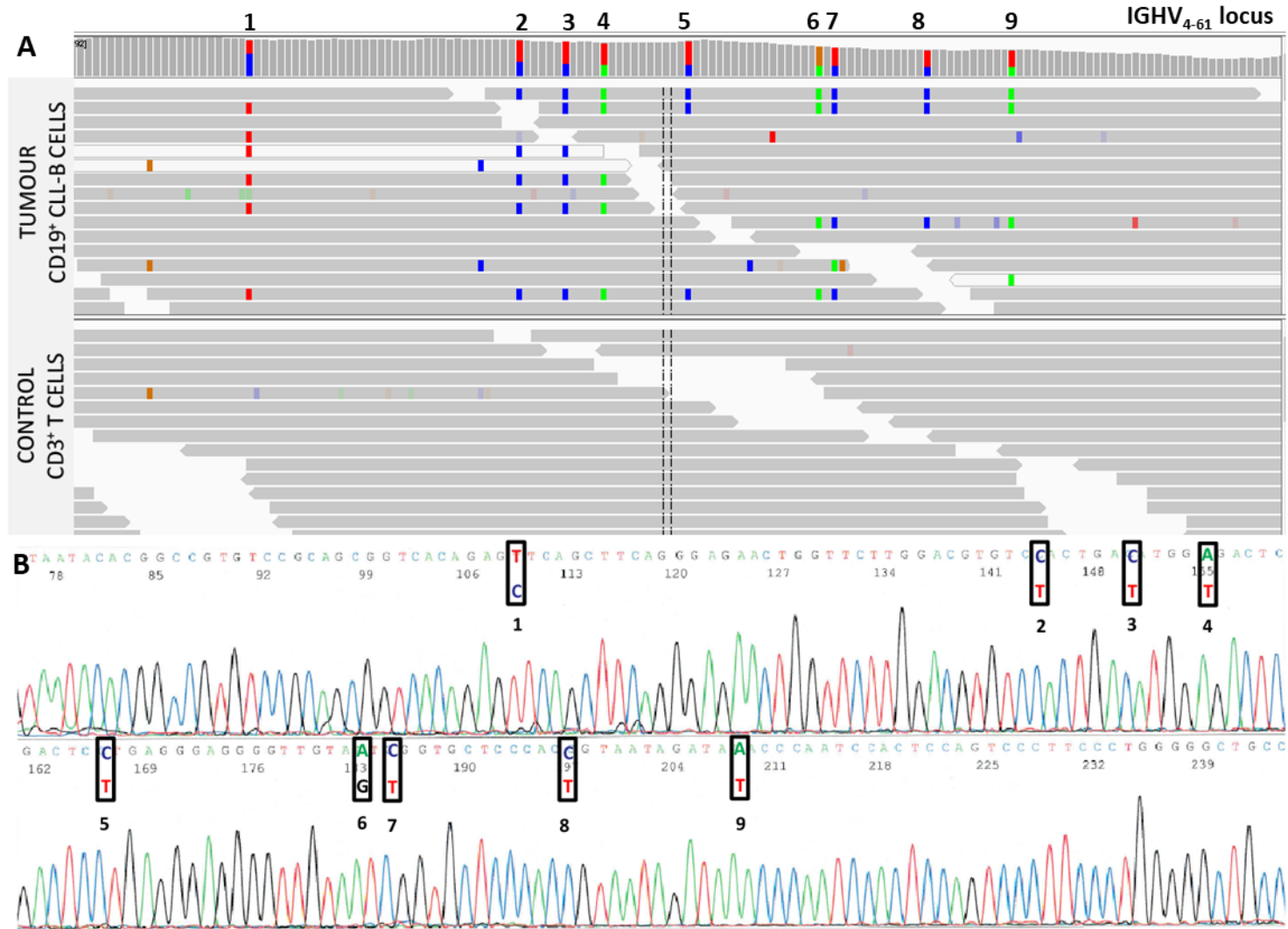


Figure 5.15. IGHV status in DB17 CLL patient sample.

Mutations present on the IGHV₄₋₆₁ locus. Chr14:107,095,164-107,095,288 shown on **(A)** IGV and **(B)** by Sanger sequencing from the CLL patient sample. The 9 mutations identified in this region are indicated. The nucleotides written within the black boxes represent the mutated (top) and the reference (bottom) allele. Colour legend: red (T), blue (C), green (A) and brown (G).

5.4.6 The common CLL 13q deletion arising from a 6q:13q translocation

Several common CNVs have been identified in CLL patients: trisomy 12 (20%), 2p gain (5-28%), and deletion of 11q (6-20%), 13q14 (50%), 17p13 (5-10%) and 6q21 (5-7%) with distinct impact on the disease. Deletion in 13q14 is found in ~50% of CLL patients and associates with a better prognosis when identified as the only abnormality. Deletion of 6q21 is present in 5-7% of patients and the prognostic significance is unknown (Ghamlouch *et al.*, 2017).

The detection of CNV using cn.MOPS from WGS of the tumour/control pair (**Supplementary Figure 4**) revealed a deletion at 6q21 (Chr6:107975001-109465000, 1.49Mb) and at 13q14.2-13q14.3 (Chr13:50525001-51500000, 0.98Mb) (**Figure 5.16A**). The number of reads covering these regions, decreased from 70 to 35, therefore suggesting that both events were clonal and mono-allelic. These results were also in agreement with data obtained by Dr Thet Lin using array CGH in a sample from the same patient (**Figure 5.16B**). The genes deleted in the 6q21 and 13q14.2-13q14.3 CNV are listed in **Supplementary Figure 5**. The 13q14 deletion included the minimal deleted region (MDR) that is found in most cases and encompasses *DLEU2*, *MIR16-1* and *MIR-15A*. Genes within the MDR regulate B cell proliferation and MDR deletion results in CLL (Klein *et al.*, 2010).

Surprisingly, observation of both regions in IGV revealed that the 6q21 and 13q14 deletions originated from a 6q:13q translocation. To further corroborate this observation, a pair of primers was designed surrounding each fusion point: A (Chr6:107,974,284-Chr13:51,501,490) and B (Chr13:50,523,174-Chr6:109,465,113). PCR and Sanger sequencing was successful across the two translocation fusion points in DB17 but not in other 8 CLL patient samples (DB59-66) (**Figure 5.16C-F**, **Supplementary Figure 6** and **Supplementary Figure 7**).

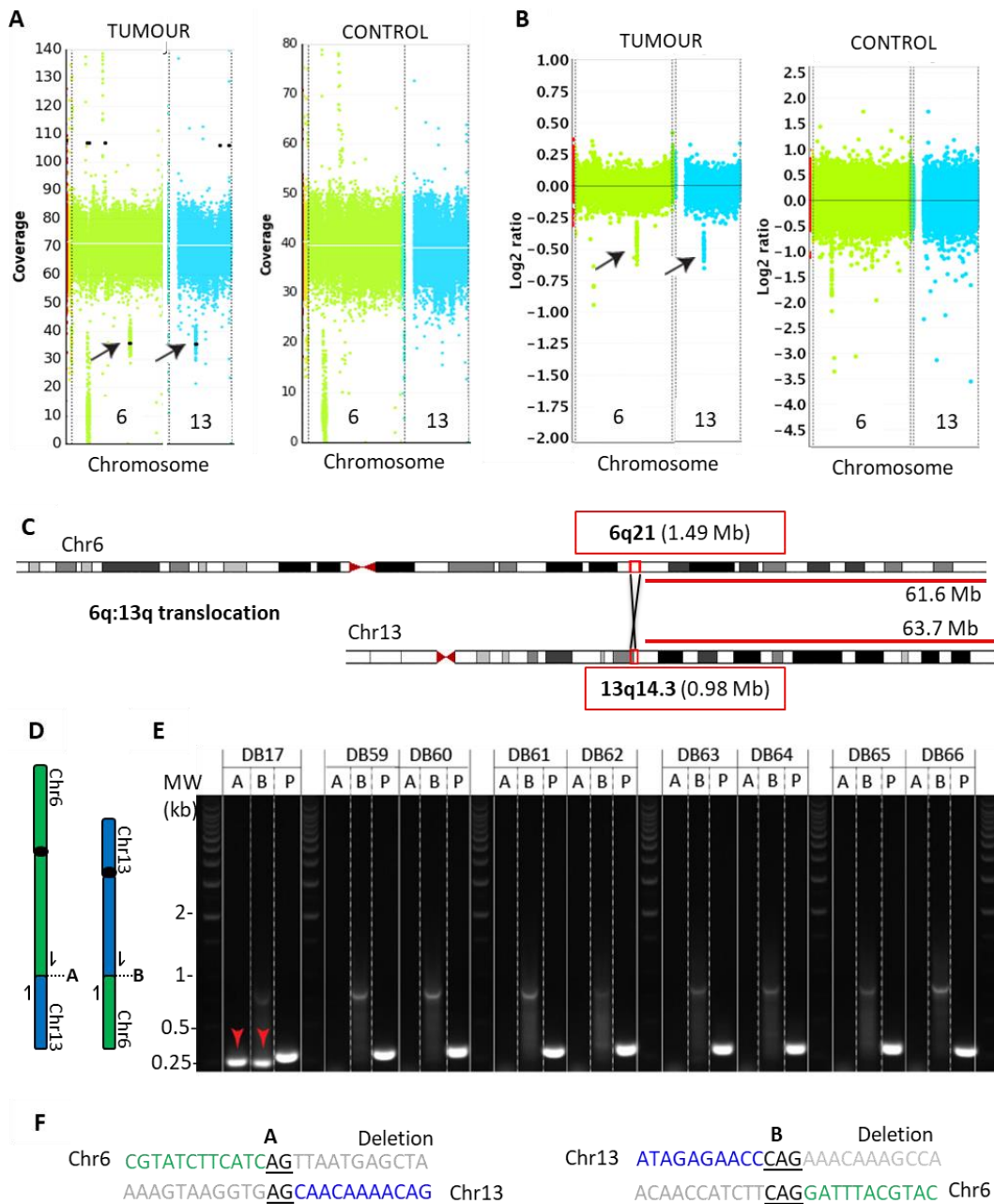


Figure 5.16. Common 13q deletion originated from 6q:13q translocation

(A) CNV observed using cn.MOPS (B) CNV observed using array CGH (provided by Dr Thet Lin) (C) Cartoon representation of the 6q:13q translocation causing deletion at Chr6q21 (1.49Mb) and Chr13q14.3 (0.98Mb). (D) Illustration of the 6q:13q translocation with pairs of primers to amplify the fusion points A (Chr6:107,974,284-Chr13:51,501,490) and B (Chr13:50,523,174-Chr6:109,465,113). (E) PCR for the validation of the 6q:13q translocation (fusion points A and B) in the WGS patient DB17 as well as the other 8 CLL patients that underwent NGS analysis of telomere fusion amplicons in Chapter 4 (negative control). As an internal control (P) reamplification of a section of the POT1 gene (primers POT2A and POT2B). All PCR products expected around 250bp. Red arrowheads indicate the presence of the translocation. (F) Sequencing of the breakpoints A and B. Deleted sequence in grey, microhomology in black underlined, Chr6 in green and Chr13 in blue.

5.4.7 Copy neutral loss of heterozygosity: 17p uniparental disomy (17p UPD) with mutated *TP53*

SNVs were not regularly distributed for T cells: 69.7% (5209/7649) of total SNVs detected in the control sample concentrated at chr17p ($0-1.765 \times 10^7$) (**Figure 5.17BC**). These results were indicative of LOH at 17p in the CLL-B cells. Although LOH usually arises from the mono-allelic loss of genetic material, a deletion was not identified using cn.MOPS or array CGH at 17p (**Supplementary Figure 4**). Altogether, results indicated the presence of a copy neutral LOH (CN-LOH) or UPD at 17p in the tumour CLL-B cells (**Figure 5.17**).

From the total number of SNVs identified in CLL-B cells, 0.44% (17/3821) located at the 17p LOH region which may have been acquired after LOH. From those, 17.6% (3/17) were polymorphisms and 82.4% (14/17) mutation. A missense mutation at Chr17:7,578,395 in the *TP53* gene was predicted to have a deleterious effect since it triggered a change of histidine 179 to tyrosine in the protein core (**Figure 5.18**).

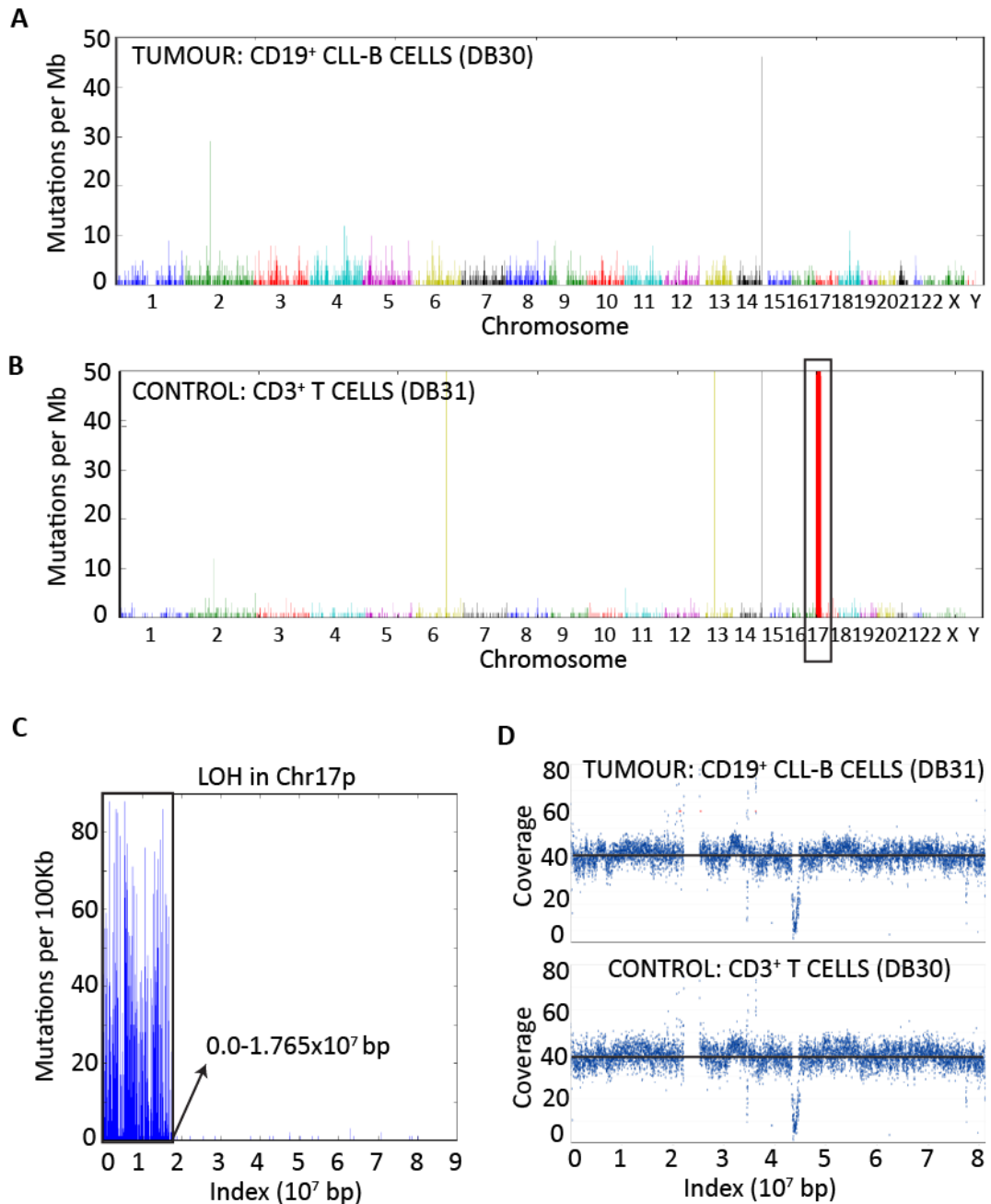


Figure 5.17. 17p Copy Neutral Loss of Heterozygosity (CN-LOH).

Distribution of mutations per Mb across the genome for **(A)** Tumour CD19⁺ CLL-B cells (DB30) and **(B)** Control CD3⁺ T cells (DB31). Region with high frequency of mutations at Chromosome 17 highlighted within black box. **(C)** Zoom into region with high frequency of mutations at Chr17:0-1.765 $\times 10^7$. **(D)** Coverage of Chr17 for both tumour and control fractions.

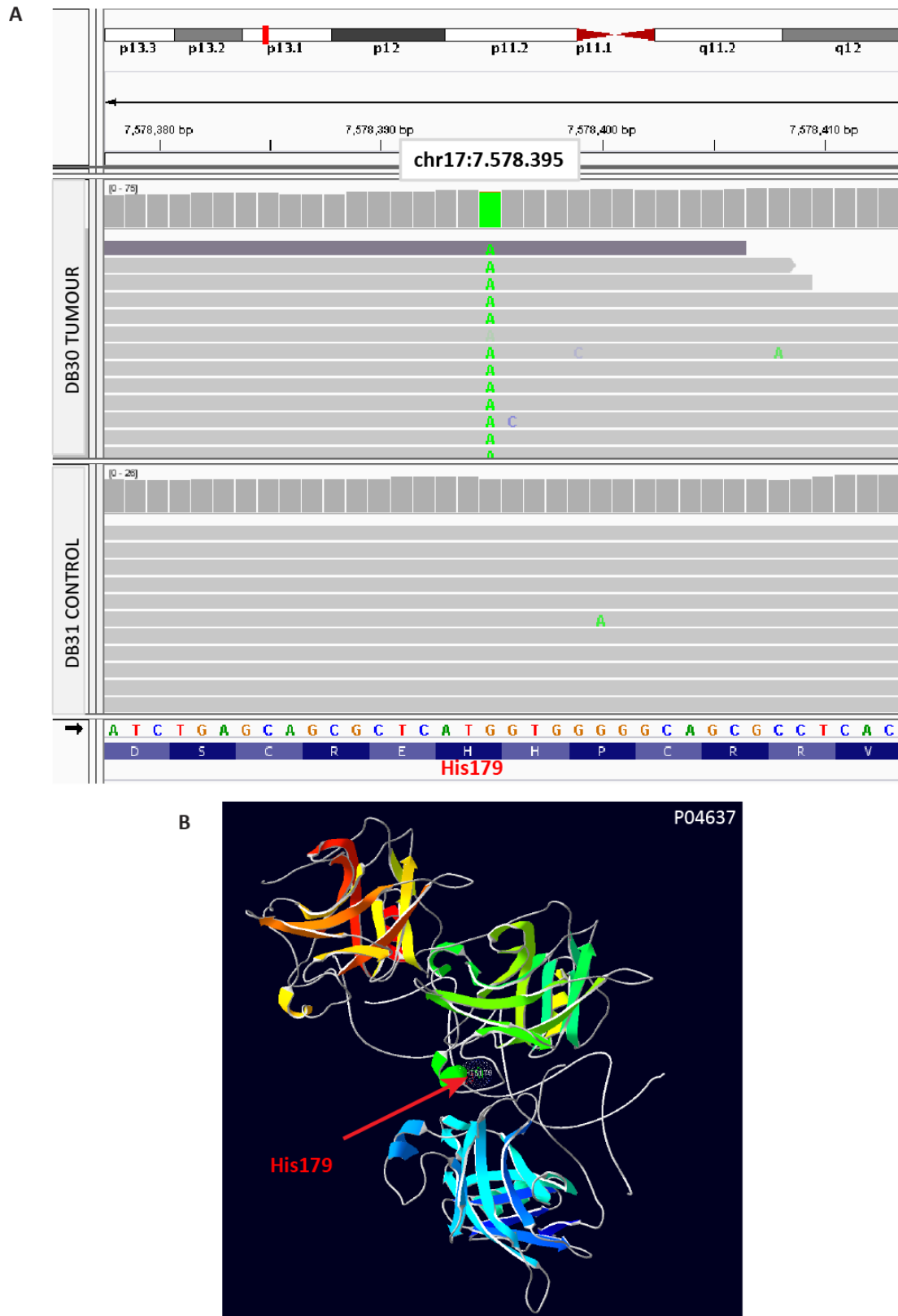


Figure 5.18. DB17 CLL patient sample with a homozygous *TP53* mutation.

(A) Observation of the mutation on IGV. **(B)** Structure of tumour protein p53 (ID: P04367) visualised on Swiss Pdb (Guex and Peitsch, 1997). Red arrow points the location of the His179Tyr mutation.

5.5 DISCUSSION

5.5.1 Multiclonality: theories

Potential biconality was observed in a CLL patient sample (DB17) with two distinct XpYp TL distributions using STELA. However, it could have also indicated a biallelic TL distribution or cellular contamination but both possibilities were excluded in subsequent experiments. The TL measurement of the patient CLL-B cells over 7 years indicated that the population with shorter telomeres declined over time (**Figure 5.2**). These results were further corroborated *in vitro* after comparing the TL profile of DB17 at baseline with day 28 (**Figure 5.6**). The possibility of distinct cellular composition was ruled out after measuring the TL of purified CD19⁺ CLL-B cells compared with CD3⁺ T cells (**Figure 5.4**). Biconality or polyclonality was subsequently verified with the VAF obtained from SNP data analysis (**Figure 5.5**). These results suggest that the TL profile obtained with STELA may, in some cases, be a potential indicator of intra-tumour heterogeneity.

There were two possible explanations for the emergence of multiple clones in this patient. The first model proposes 2 distinct clones stably maintained over time: one with short and another with long telomeres. However, this hypothesis was not likely since the WGS data revealed that most mutations were clonal and present in all cells with only a small percentage that were subclonal. Therefore, the second hypothesis was most probable: a CLL first clone with longer telomeres that became more heterogeneous over time, from which a subclone originated and expanded until reaching equilibrium while its telomeres continued to erode (**Figure 5.19**).

To more formally examine this hypothesis, it would be important to isolate both clones in culture to characterise their TL and frequency of fusions in order to assess whether telomere instability is exclusive to the subclone with shorter telomeres. Treatment is known to change the tumour composition and select for the fittest clones which usually associate with relapse (Landau *et al.*, 2013). Since the CLL patient has never been treated it is most likely that the distinct clones are found in equilibrium. In the future it would be important to perform *in vitro* experiments that would allow comparing the responses of each clone to selected CLL chemotherapeutics, particularly if the patient needs to be treated.

Equilibrium of distinct subclones is generally associated with better outcome compared with competition (Landau *et al.*, 2014). This chapter's original hypothesis stated that the

increased frequency of fusions was product of the subclone with short telomeres. However, the patient's stable disease could be partially explained by the equilibrium with the clone with longer telomeres.

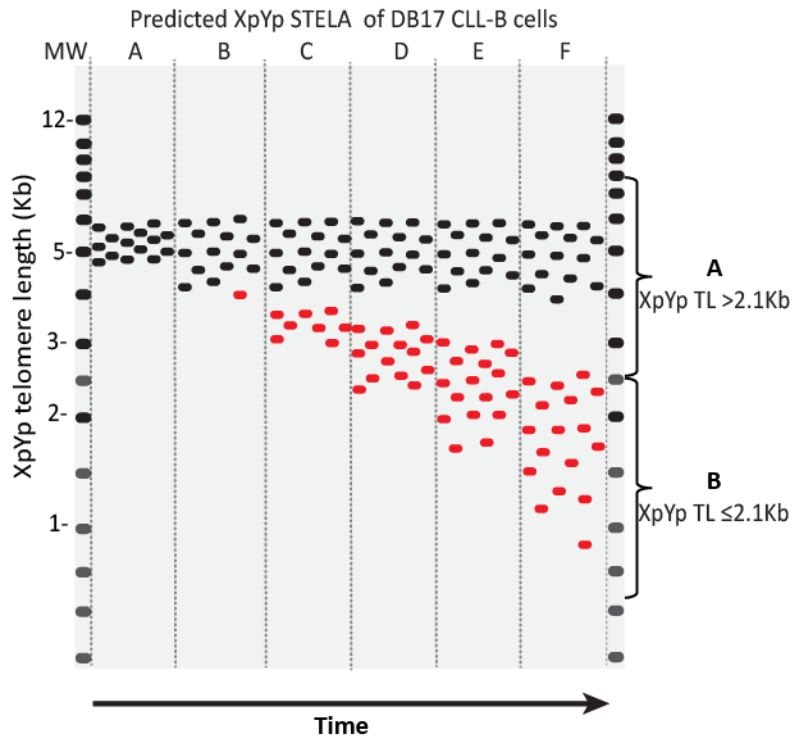


Figure 5.19. Predicted evolution of the XpYp TL for the patient CLL-B cells.

Cartoon representation of the predicted TL evolution observed with STELA. A homogeneous TL distribution becomes more heterogeneous. A subclone with shorter telomeres (red) expands but its telomeres erode over time. Both populations finally reach the bimodal distribution.

5.5.2 Impact of mutations

5.5.2.1 CLL drivers and the MAPK/ERK signalling pathway

Enrichment analysis from the list of mutated genes identified the RAS-RAF-MEK-ERK (MAPK/ERK) signalling pathway. This route culminates in controlling cell growth, cell survival, metabolism, invasion and senescence (Dhillon *et al.*, 2007). The MAPK/ERK pathway has also been shown upregulated in a subset of CLL patients by transcriptomic analysis (Ferreira *et al.*, 2014). It is possible that the genes involved in the MAPK/ERK pathway were actively transcribed in the CLL-B cells and thus exposed to DNA damage. Targeted genes contained clonal mutations present in all cells, suggesting it might have been an early event.

Amongst the mutated genes, *TP53*, *POT1*, *CREBBP*, *BRAF* and *ATR* have been identified in CLL and are a part of the DNA damage response, apoptosis, NF- κ B signalling, cell cycle, NOTCH1 signalling and B-cell signalling (Puente *et al.*, 2015, Guieze and Wu, 2015) (**Figure**

5.10). The first three were clonal mutations while mutations in *BRAF* and *ATR* were subclonal.

A novel mutation was detected in the epigenetic regulator cAMP-response element binding protein (*CREBBP*) predicted to cause an amino acid substitution (p.Asn1131Thr) with damaging effect. *CREBBP* is involved in several functions including cell cycle, antigen presentation and repair of DNA damage (Kretsovali *et al.*, 1998, Chae *et al.*, 2015). *CREBBP* regulates the transcription of the CLL driver *NOTCH1* and may modulate the MAPK/ERK signalling pathway (Bienvenu *et al.*, 2010, Dixon *et al.*, 2017). Mutations in this gene are present in many cancers, particularly contributing to lymphomagenesis in germinal-centre B-cells (Zhang *et al.*, 2017). Other novel clonal missense mutations triggering an amino acid substitution with predicted deleterious effect that may be relevant to the pathogenesis of the disease include: *AIFM1* (p.Pro545Arg) and *CD83* (p.Tyr45His). The apoptosis-inducing factor, mitochondrion-associated, 1 (*AIFM1*) plays a role in apoptosis and has been associated with tumorigenesis (Joza *et al.*, 2001). Interestingly, the CD83 antigen (*CD83*) is expressed by activated B lymphocytes and has a role as an immunosuppressive agent and a regulator of lymphocyte survival and differentiation (Kretschmer *et al.*, 2009). In addition, elevated levels of soluble CD83 in plasma of patients with CLL have been associated with shorter treatment-free survival (Hock *et al.*, 2009).

In addition to *BRAF* and *ATR*, subclonal mutations were also identified in *MEIS1*, *IGF1R*, *RET* and *MYB* amongst other. None of them was predicted to produce a change at protein level; however, mutations in regulatory sequence can also modify gene expression. Ataxia Telangiectasia and Rad3 related (*ATR*) protein kinase is important for preventing telomere fragility and telomere fusions, and the recruitment of telomerase (McNees *et al.*, 2010, Tong *et al.*, 2015). The oncogene *MYB* and the insulin-like growth factor-1 receptor (*IGF1R*) are overexpressed in some cases with CLL. In addition *IGF1R* been identified as a therapeutic target in CLL (Vargova *et al.*, 2011, Yaktapour *et al.*, 2013). Both clonal and subclonal mutations were identified in *MEIS1*, a transcription factor required for self-renewal in normal and leukemic haematopoiesis and suggested to activate the PI3K/Akt and MAP kinase signalling pathways (Wong *et al.*, 2007, Heuser *et al.*, 2011, Gibbs *et al.*, 2012). In addition, higher levels of MEIS are associated with resistant Acute Myeloid Leukaemia (AML) and it has been identified as a promising prognostic marker (Liu *et al.*, 2017a, Mohr *et al.*, 2017).

These data must be interpreted with caution since it is not possible to distinguish driver from passenger mutations merely from the genetic data. A further study with more focus

on downstream gene expression analysis, to accurately assess the impact these mutations have in the cell, is therefore suggested.

5.5.2.2 Truncated REV3L and chromosome instability

A novel mutation in the *REV3L* gene predicted to create a truncated protein (1426/3130 amino acids) was identified from this dataset (**Figure 5.20**). *REV3L* is the catalytic subunit of the low fidelity DNA polymerase zeta (Pol ζ) which is implicated in the tolerance of DNA damage in translesion synthesis (TLS) and is important for maintaining genome stability (Lange *et al.*, 2013). *REV3L* interacts with *REV7* (*MAD2L2*), which confers resistance to DNA damage (cisplatin and UV radiation) preventing DNA breaks and maintaining chromosome stability (Tomida *et al.*, 2015). This *REV3L* p.Gln1426* novel mutation would cause the loss of the *REV7* binding domain and the polymerase domain, leading to a dysfunctional protein. Moreover, depletion of *REV3L* results in more common fragile site expression, chromosome breaks and increased anaphase bridges (Bhat *et al.*, 2013).

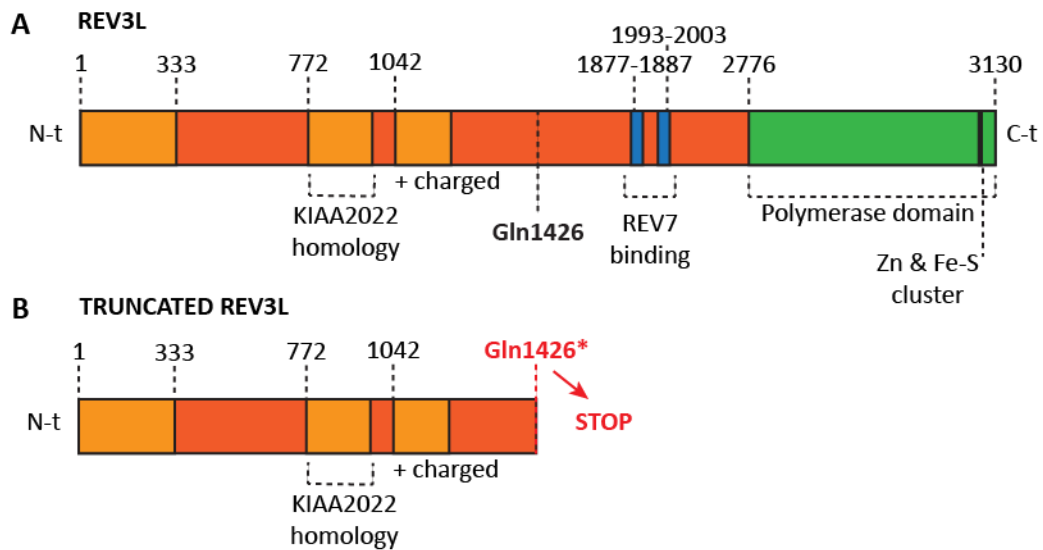


Figure 5.20. *REV3L* p.Gln1426* mutation causes loss of polymerase domain and *REV7* binding site.

Primary structure of the *REV3L* protein (UniProt ID: O60673), the catalytic component of the polymerase ζ . (A) Intact and (B) truncated protein caused by a somatic mutation (p.Gln1426*) coding for a stop codon identified in the CLL patient DB17, that provokes the loss of the *REV7* binding domain (blue) and the polymerase domains (green). Adapted from Lange *et al.*, 2016.

REV3L is as a promising therapeutic target in several cancer types and its overexpression conferred resistance to DNA damaging agents allowing a higher yield of viable cells. In contrast, its depletion resulted in a high incidence of chromosome breaks leading to cell

death and sensitised the cells to chemotherapeutic agents like cisplatin (Lee *et al.*, 2014b, Yang *et al.*, 2015, Lange *et al.*, 2016, Huang *et al.*, 2016).

It appears that this clonal heterozygous mutation in *REV3L* creates a truncated protein missing the REV7 and POLZ binding domains, which may explain the high frequency of telomere fusions in this patient. Further research is required to investigate the effect of this high-impact *REV3L* mutation, on telomere dysfunction. This could be undertaken by modelling the Gln1426* mutation in the CLL cell line MEC1 (Stacchini *et al.*, 1999) using genome editing techniques, following measurement of the TL using STELA (Baird *et al.*, 2003) and telomere fusion activity (Capper *et al.*, 2007).

This discovery may provide key information for the pathogenesis of this patient. Intact Pol ζ provides the cell with survival advantage as it allows TSL replication past DNA lesions. However, dysfunctional *REV3L* may result in increased collapsed replication forks and it is associated with increased sensitivity to cisplatin and higher chromosomal aberrations. In this case, the mutation in *REV3L* may contribute to the patient's indolent disease by negatively impacting on the viability of the CLL cells. Therefore, it will also be important to assess whether *REV3L* can be a therapeutic target in CLL. This could be assessed by establishing CLL cell lines with *REV3L* suppression or overexpression and determine their proliferative impact, apoptosis rate and response to cisplatin and other chemotherapies.

5.5.3 POT1 and telomeres

POT1 is a component of the shelterin complex that binds telomeric DNA, regulating its accessibility to DNA-modifying enzymes (Baumann and Cech, 2001). Mutated *POT1* is a CLL driver. Mutations in the OB domain of this gene that disrupt the POT1-DNA binding site, have been associated with telomere dysfunction in CLL (Ramsay *et al.*, 2013).

Disruption of the POT1-TPP1 interaction was shown to decrease the binding affinity of POT1 to DNA (Rice *et al.*, 2017). In this study the presence of 7 mutations predicted to disrupt that interaction was investigated in 31 CLL patients with short telomeres (**Figure 5.11**). However, these mutations did not appear to be enriched in CLL patients with short telomeres.

Mutations in the OB domain disrupting the interaction with DNA have been identified in many cancers and shown to promote genomic instability by initiating the DNA damage response (A-NHEJ) at telomeres that leads to chromosome fusions (Gu *et al.*, 2017). POT1 forms a complex with TPP1 to regulate TL by controlling the access of telomerase (Tejera *et*

al., 2010). It has been observed that POT1 mutations in the OB domain or defective POT1-TPP1 complex result in defective elongated telomeres (Gu *et al.*, 2017, Rice *et al.*, 2017). Therefore it is conceivable that mutations disrupting the POT1-TPP1 heterodimer may result in telomere fusion in CLL-B cells that exhibit long telomeres. Investigation of these mutations and association with dysfunctional telomeres is required in a cohort of CLL patients with long telomeres.

Whole genome analysis of DB17, the CLL patient with high fusion frequency in this study, revealed a novel clonal mutation in *POT1* antisense RNA 1 (*POT1-AS1*) that may contribute to the genomic instability of this patient. Functional studies by modelling this new mutation in the CLL cell line MEC1 (Stacchini *et al.*, 1999) using gene editing technology (CRISPR/CAS9-mediated mutation) will enable understanding the role and impact of this *POT1-AS1* novel mutation compared to other CLL recurrent *POT1* mutations in TL, stability and fusion activity.

5.5.4 Mutated CLL and a signature of mutations consistent with nc-AID

The *IGHV* mutational status has been known as an important prognostic marker in CLL for almost two decades (Hamblin *et al.*, 1999, Damle *et al.*, 1999). In this case study, the patient presented mutated *IGHV* CLL (**Figure 5.15**) which is generally associated with indolent disease and is most likely originated from a memory B-cell that have undergone *IGHV* somatic hypermutation at the germinal centre (Seifert *et al.*, 2012).

The frequency of somatic mutations ranges from 0.001-400 mutations/Mb depending on the cancer type (Alexandrov *et al.*, 2013a). In this study, patient DB17 CLL-B cells had a mutational frequency of 1.08 mutations/Mb, with a total of 3,607 point mutations identified across the genome. These results are in line with previous observations in CLL. Kasar *et al.*, identified an average of 0.92 mutations/Mb (3,055 mutations genome-wide) from 30 CLL patients (Kasar *et al.*, 2015). Similarly, Puente *et al.*, identified an average of 0.87 and 0.89 mutations/Mb for CLL and MBL, respectively (Puente *et al.*, 2015).

The pattern of somatic mutations identified for the CLL patient sample corresponded to Alexandrov's signature 9 (**Figure 5.12**) present in cancer cells undergoing immunoglobulin gene hypermutation (Alexandrov *et al.*, 2013a). Somatic hypermutation is exclusive to B-lymphocytes at immunoglobulin regions and is mediated by canonical AID (c-AID) (Arakawa *et al.*, 2002). However, c-AID creates a distinct signature of mutations (clustered C>T/G, signature 2) not accordant with the pattern observed in signature 9 (non-clustered T>G)

(**Figure 5.13**) (Alexandrov *et al.*, 2013a). In a later study, Kasar and colleagues proposed a role for non-canonical AID (nc-AID) earlier in tumour development in CLL patients with signature 9 characterised with non-clustered mutations (Kasar *et al.*, 2015).

Off-target activity of AID/APOBEC deaminases at unrepaired single-strand breaks (SSBs) during S or G2/M phase can result in deamination that has been associated with kataegis in cancer (Alexandrov *et al.*, 2013a, Rebhandl *et al.*, 2014, Casellas *et al.*, 2016). However, for DB17, localised hypermutation (kataegis) was exclusive to *IGLK* (Chr2p11.2) and *IGHV* (Chr14q32.33) immunoglobulin loci (**Figure 5.13** and **Figure 5.14**). In addition, the mutations identified from DB17 CLL-B cells were not clustered instead they were spread across the genome, consistent with nc-AID. Through the low-fidelity non-canonical DNA mismatch repair pathway, nc-AID uses the error-prone polymerase η contributing to mutagenesis (Pena-Diaz *et al.*, 2012). The author also proposed that nc-AID may also be responsible for inducing DSBs, translocations and telomere fusions in B cells.

Altogether, it is possible that nc-AID may have contributed to the somatic mutations present in this patient's tumour, playing a role in progression of the normal B-cell to a malignant CLL cell in this patient.

5.5.5 6q:13q translocation leads to 13q14 and 6q21 deletion in CLL patient.

For this patient, clonal monoallelic deletion at 13q14, including the minimal deleted region (MDR), and 6q21 were identified.

The prognostic significance of 13q14 deletions is debated in the literature. The 13q14 deletion is the most common chromosomal aberration in CLL (50-60%) and as a sole abnormality, has typically been associated with an indolent disease (Dohner *et al.*, 2000). The deleted section at 13q14 sometimes can include the MDR that harbours *DLEU2*, *MIR16-1* and *MIR-15A*, which are considered to have a tumour suppressor function in CLL. They are negative regulators of *BCL2* expression, regulating cell cycle and apoptosis, and therefore the loss of the MDR may result in anti-apoptotic resistance and is associated with a more aggressive disease (Klein *et al.*, 2010). However, later studies in CLL patients showed that neither monoallelic nor biallelic 13q deletion, in the absence or presence of MDR, were sufficient to be considered an adverse prognostic factor (Puiggros *et al.*, 2013, Grygalewicz *et al.*, 2016).

Deletions in 6q (including 6q21) were observed in 5 to 7% of CLL patients (dependent on the study) and were associated with a higher lymphocyte count, CD38 positivity and an

intermediate-risk prognosis (Stilgenbauer *et al.*, 1999, Cuneo *et al.*, 2004). 6q21 deletion has been linked to the tumour suppressor *ZNF292*, a growth hormone-dependent transcription factor (Puente *et al.*, 2015, Ghamlouch *et al.*, 2017).

Surprisingly, the current study in DB17 found that the 13q14 and 6q21 deletion originated from a 6q:13q unbalanced inter-chromosomal translocation (**Figure 5.16**). This was not an isolated case as it was reported that about 37% of cases with 13q deletions occurred by an inter-chromosomal unbalanced translocations (Hruba *et al.*, 2012, Kasar *et al.*, 2015). In this case, the partner was chromosome 6q. CLL patients with 13q14 deletion arising from 13q translocations have been associated with patients with more unstable genome and poor clinical outcome (Puiggros *et al.*, 2014).

In this patient, the monoallelic 6q:13q unbalanced translocation could have been an early event in the initiation of malignancy since it was present in all clones.

5.5.6 17p UPD, *TP53* and telomeres

Copy neutral loss of heterozygosity (CN-LOH) or uniparental disomy (UPD) was identified at the 17p chromosome in the patient's tumour CLL-B cell fraction. This chromosomal aberration resulted from the loss of a portion of chromosome 17p and replacement by the duplication of the remaining allele (paternal or maternal). This resulted in the loss of the polymorphic differences that existed between the two alleles, giving place to two exact copies of genetic information for that specific segment of the chromosome, including a homozygous mutation in *TP53* (**Figure 5.17** and **Figure 5.18**).

Unlike inherited UPD, the creation of somatic or acquired UPD is not well understood. Firstly, it may arise from an attempt to correct for an initial deletion using the remaining fragment as a template for duplication, or secondly as a result of mitotic homologous recombination with gene conversion (Stephens *et al.*, 2006). Locations closer to the centromeres or telomeres have been identified as hotspots for mitotic recombination that lead to UPD (Stephens *et al.*, 2006). UPD can contribute to clonal outgrowth and it presents a mechanistic role in cancer including myeloid and lymphoid malignancies such as CLL. Some of the regions affected include genes relevant to the pathogenesis of the disease such as UPD2q (*MAP2* deletion in MCL), UPD5q (*APC* mutation in colorectal cancer), UPD9p (*JAK2* mutation in AML), UPD13q (*BRCA2* mutation in ovarian cancer, and miR-15a and miR-16-1 deletion in CLL), UPD17p (*TP53* mutation in MDS and CLL) (Jasek *et al.*, 2010, Makishima and Maciejewski, 2011).

In this study, 17p UPD was identified in a CLL patient with the highest frequency of telomere fusions. Moreover, short telomeres have been associated with UPD in CLL (Sellmann *et al.*, 2016), therefore it is possible to hypothesise that telomere dysfunction and cycles of breakage-fusion-bridge may drive UPD in cancer.

In addition, the patient's CLL-B cells presented a clonal *TP53* mutation that might explain the increased frequency of telomere fusions observed. Besides, it has been reported that CLL patients with mutated *TP53* present high telomerase expression, short fusogenic telomeres, high levels of genomic instability and associate with poor outcome (Gueze *et al.*, 2016). Surprisingly and contrary to expectations, the CLL patient presents an indolent CLL and has been treatment-free for 14 years.

In the case study exposed in this Chapter, the following model was proposed to explain the development of the 17p-UPD. Initially, an initial heterozygous *TP53* clonal mutation (p.His179Tyr) (primary event) partially inactivates p53, inhibiting the cellular responses to dysfunctional telomeres. This may facilitate telomere instability and the initiation of cycles of fusion and breakage of chromosomes. Then, intra/inter-chromosomal fusion with 17p telomere results in the loss of the 17p arm containing the protective wild type allele (17p LOH). Subsequently, a duplication of the chromosome segment (17p UPD) results in the homozygous *TP53* mutation that inactivates the tumour suppressor gene which permits further cell proliferative advantage, resulting in clonal outgrowth (**Figure 5.21**). Interestingly, the mutation was present in 95% of the reads in the dataset. It was therefore possible to hypothesise that it could be a primary and clonal event implicated in the initiation of malignancy.

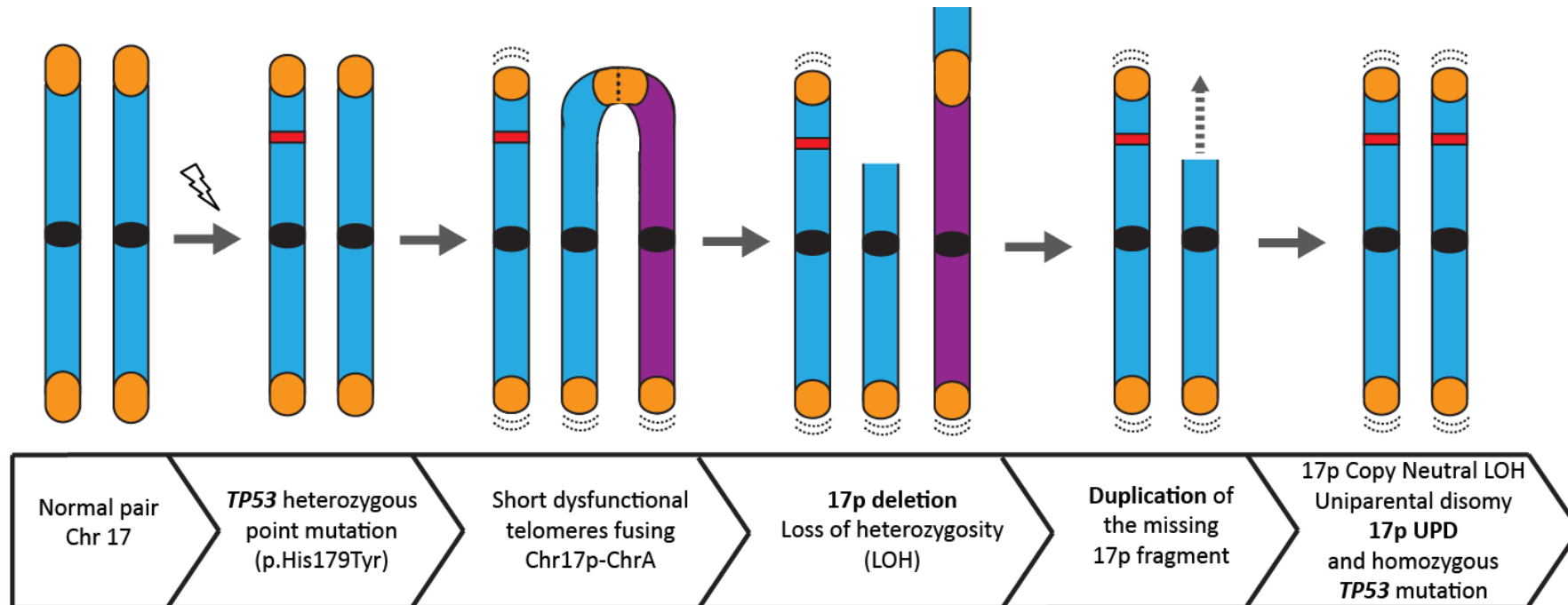


Figure 5.21. Proposed model explaining the origin of the 17p UPD in the patient CLL-B cells.

Clonal event present in all cells. Normal pair of chromosome 17 undergoes an initial mutational event (white lightning bolt): a heterozygous *TP53* somatic missense mutation (red) that triggers an amino acid substitution (p.His179Tyr). This event was followed by the initiation of telomere instability and potential cycles of BFB. In this case a 17p telomere fuses with another telomere (ChrA) leading to a 17p deletion ($0-1.765 \times 10^7$) of the allele containing the WT *TP53*, and loss of heterozygosity (LOH). The remaining 17p arm was duplicated leading to two exact copies of the 17p fragment, including the homozygous *TP53* mutation. Finally, 17p copy neutral LOH or 17p uniparental disomy (17p UPD) with homozygous *TP53* mutation was observed from the WGS.

5.5.7 Excessive genetic instability may negatively impact on cancer cell fitness resulting in patient's indolent disease

The biclonal CLL patient investigated in this case study presented the accumulation of many adverse prognostic factors: from driver mutations to telomere dysfunction and chromosomal rearrangements (**Figure 5.22**). Contrary to expectations this patient presented a mild and asymptomatic disease and has been treatment-free for 14 years.

Given the results obtained from the WGS of DB17 tumour/control, a timeline of events contributing to CLL in this particular patient is proposed. An early clonal event may have been a mutated *TP53*, which allowed the initiation of telomere fusions and may explain the increased frequency fusions detected in this patient. Cycles of BFB could have driven large-scale genomic rearrangements including 17p UPD. Additionally, other chromosomal aberrations like the 6q:13q translocation may have occurred.

In parallel, a non-canonical DNA mismatch repair driven by nc-AID and Pol η may have contributed to the majority of somatic mutations present in the CLL-B cells. In addition, a potentially dysfunctional Pol ζ might have resulted in decreased tolerance to TLS DNA damage. This is characterised with the accumulation of replication forks, anaphase bridges and chromosomal breaks that negatively impact cellular fitness, partially explaining the increased frequency of fusions. Most important, by preventing the accumulation of cells with acquired mutations it most likely contributed to the patient's indolent disease. Further research is required; however, REV3L, the catalytic subunit of Pol ζ may be a potential therapeutic target in CLL sensitising cells to DNA damage.

Altogether, this is in line with Birkbak and colleagues' discovery who proposed that an excess of chromosomal instability (CIN) is poorly tolerated by the cancer cells, resulting in a better prognosis for the patient (Birkbak *et al.*, 2011). For this particular CLL patient, this scenario may be the case.

Interestingly, the presence of intra-tumour heterogeneity has been confirmed. Thus, it is proposed that later during malignant progression, a subclone with short telomeres may appear from the primary clonal population. The telomeres of the subclonal population may continue to erode, as it was observed from the CLL patient samples obtained over 7 years, as well as *in vitro*.

The distinct clones may be in equilibrium; however, if in the future the patient needed treatment, it is important to consider that therapy may change the tumour composition.

This would allow the fittest clone to survive, potentially resulting in relapse. Therefore, while the patient presents an asymptomatic disease (stage A) and given the tumour's genetic information, the patient is better untreated. However, if therapeutic options need to be considered in the future, it is important to reflect that the standard of care FCR treatment would be damaging for the patient given the *TP53* mutation associated with high-risk CLL (Tam and Stilgenbauer, 2015). In contrast, dysfunctional Pol ζ provides increased sensitive to cisplatin (Lee *et al.*, 2014b, Huang *et al.*, 2016). Other therapeutic options like Ibrutinib or Venetoclax should be considered.

This chapter has improved the understanding of the disease of a CLL patient and proposed a timeline of events contributing to malignancy including the evolution of intra-tumour heterogeneity. This work has emphasised the importance of telomere instability which may drive chromosomal rearrangements and it has identified novel mutations that may be relevant to telomere instability, CLL pathogenesis and an indolent disease.

Altogether, this chapter highlights the potential of precision medicine to establish prognosis and elaborate patient-tailored treatment.

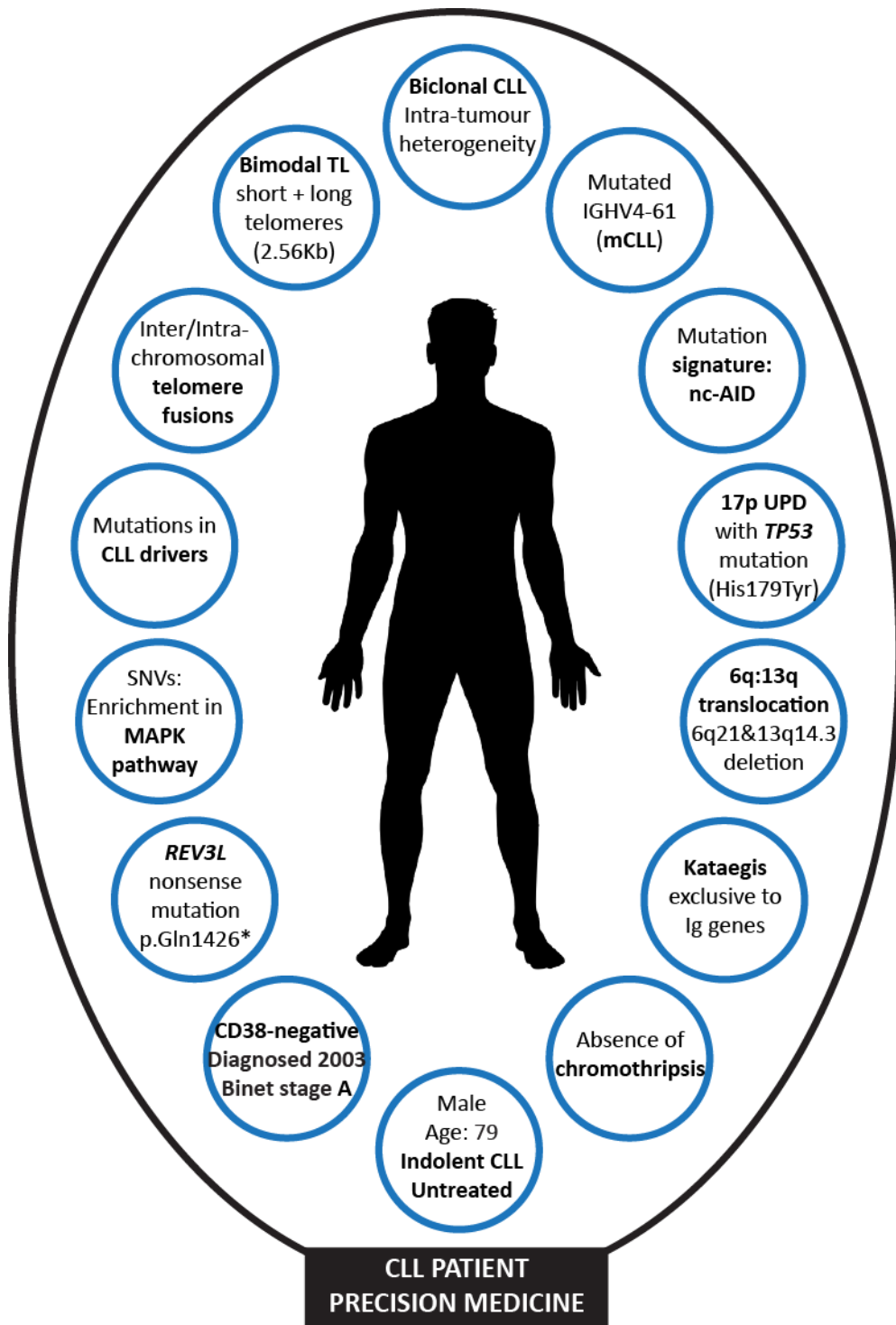


Figure 5.22. The paradoxical effect of excessive genetic instability in cancer progression.

Integrated study summarising the information acquired for the CLL patient. Telomere length distribution, frequency of telomere fusions, intra-tumour heterogeneity, molecular markers, SNVs, cytogenetics and other information relevant to CLL is reported.

CHAPTER 6:

GENERAL DISCUSSION AND FUTURE DIRECTIONS

6.1 SUMMARY

This study aimed to assess the impact of telomere dysfunction on the cancer genome of CLL patients, with a particular focus on the 5p telomere that is proximal to the *hTERT* locus.

The aim of chapter 3 was to characterise the 5p telomere. STELA, TVR-PCR and the telomere fusion assay were successfully adapted to study the 5p chromosome end. The 5p telomere length (TL) of 57 patient CLL-B lymphocytes was measured and compared with the 17p and XpYp telomeres. Although the 5p telomere initially measured significantly longer than that of XpYp (but not 17p), this difference disappeared after subtracting the length contribution of the telomere variant repeats (TVRs), indicating that the canonical TTAGGG repeat content of the 5p telomere did not differ from the other telomeres analysed.

Chapter 3 aimed to detect telomere fusions in a cohort of CLL patients with short telomeres, to identify those patients with the highest fusion frequency and to assess whether the frequency of fusions could further stratify CLL patients into prognostic subgroups. In addition, it aimed to investigate whether the 5p telomere was dysfunctional and fusogenic in CLL patients and during a telomere-driven crisis in culture. Telomere fusions, targeting a total of 24 distinct telomeres, were investigated in 276 untreated CLL patient samples with predominantly short TL (<3.81Kb) obtained from UHW in Cardiff, as well as the LRF CLL4 clinical trial and the ARCTIC and ADMIRE clinical trials. Telomere fusions were detected in 71.7% of patients with varied frequencies. The frequency of fusions did not correlate with TL and failed to significantly provide prognostic information (although shorter survival was observed among patients with the highest fusion frequency). Telomere fusions involving the 5p chromosome end were detected in 22.6% of CLL patient samples. Moreover, 5p fusions were detected in crisis-stage HCT116^{DN-hTERT} cells prior to the amplification of the 5p chromosome end and the upregulation of telomerase activity.

The aim of chapter 4 was to provide a detailed characterisation of telomere fusion amplicons, particularly those involving 5p, from CLL patients with the highest frequency of fusions. The purpose was to investigate which areas of the genome become incorporated

into telomere fusions and whether there is an association with CLL or other oncogenic pathways. In addition, it aimed to examine microhomology at the fusion junction to elucidate the DNA repair mechanisms involved. A large-scale sequence characterisation of 914 single-molecule amplified telomere fusion events detected in CLL-B cells from 9 patients with the highest fusion frequency ($>4.20 \times 10^{-5}$ per diploid genome) was undertaken using high-throughput paired-end Illumina sequencing. Head-to-head, intra- and inter-chromosomal telomere fusions were identified. It is important to note that some fusion events incorporated non-telomeric loci including protein-coding genes expressed in CLL-B cells, other oncogenes, the ancestral telomere at Chr2q13 and mitochondrial DNA. Moreover, loci with previously-reported copy number aberrations in CLL were found to be incorporated into telomere fusions. At least 19% (172/914) telomere fusions included the 5p chromosome end. 5p sister chromatid fusions that could impact telomerase expression were also detected. Characterisation of the fusion junctions revealed differential usage of microhomology associated with particular types of fusions, implicating the involvement of distinct DNA repair mechanisms including classical non-homologous end joining (C-NHEJ), alternative-NHEJ (A-NHEJ) and single-strand annealing (SSA).

Surprisingly, bimodal TL distributions were detected in 67% (6/9) of patients with the highest frequency of telomere fusions. This proportion was notably elevated when compared with the 4% (11/276) of CLL patients with bimodal TL from the ARCTIC and ADMIRE cohorts (XpYp TL range: 0.80-7.49kb) as observed by Norris *et al.*, (manuscript in preparation). The aim of chapter 5 was to investigate whether the bimodality observed using STELA was a reflection of distinct subclonal populations. Therefore, intra-patient tumour heterogeneity was investigated in one of the 9 CLL patients with the highest frequency of fusions. The presence of tumour heterogeneity was confirmed by the allele frequency of SNP data obtained after WGS of patient-matched tumour/control samples.

In addition, chapter 5 aimed to reveal the impact of the distinct clonal populations on the patient's prognosis and to examine the patient's whole genome to investigate the cause of the increased fusion frequency and the patient's indolent disease. However, most mutations and genomic rearrangements detected were clonal events, except for some low-modifier impact mutations. The patient's genome was characterised with 6q and 13q deletions arising from a 6q:13q translocation. Surprisingly, 17p UPD was identified, harbouring a homozygous *TP53* (p.His179Tyr) mutation that is associated with high-risk CLL and elevated frequency of telomere fusions. Mutations were detected in CLL driver genes and in genes involved in the MAPK/ERK pathway. In addition, a novel mutation in the

catalytic subunit of the polymerase ζ gene *REV3L* involved in DNA translesion synthesis was revealed. Altogether, potential CLL high-risk factors including increased frequency of fusions, elevated genomic instability and tumour heterogeneity were detected in a CLL patient with indolent disease (over 14 years of study to date).

6.2 GENERAL DISCUSSION

6.2.1 Differences in telomere length at distinct chromosome ends

It has been proposed that specific chromosomes have different TLs that may associate with the size of the chromosome arm (Martens *et al.*, 1998, Deng *et al.*, 2004, Wise *et al.*, 2009). Heterogeneous or multimodal chromosome-specific TL distributions in non-purified samples could be a reflection of mixed cell populations. In addition, bimodal TL distributions have been observed from the same pure cellular population, indicative of distinct allelic distribution (Deng and Lucas, 1999, Baird *et al.*, 2003, Londono-Vallejo *et al.*, 2001, Graakjaer *et al.*, 2003). However, bimodal distributions can also reflect patient intra-tumour heterogeneity (Lin *et al.*, 2010, Lin *et al.*, 2014), as demonstrated in this study. Another potential explanation for chromosome-specific differences in TL distributions may relate to TVRs, non-functional repeats located at the proximal regions of the telomere. TVRs have also been found interspersed throughout the telomere in ALT cells (Mendez-Bermudez *et al.*, 2009, Conomos *et al.*, 2012, Lee *et al.*, 2014a). The study of variant repeats is important since these sequences can disrupt shelterin binding sites and alter the formation of telomere quadruplex structures, affecting chromosome integrity (Mendez-Bermudez *et al.*, 2009, Broccoli *et al.*, 1997a, Broccoli *et al.*, 1997b). Several variants have previously been identified, including TCAGGG, TGAGGG and TTGGGG most commonly, as well as the lower frequency variants GTAGGG, TTAGGG, TTCGGG, TTTAGGG, CTAGGG, TAAGGG, ATAGGG, TTAAGGG, TTTGGG, GTGGGG, TTAGCG and TGGGGG (Allshire *et al.*, 1989, Baird, 1995, Baird *et al.*, 2003, Letsolo *et al.*, 2010, Lee *et al.*, 2014a). By high-resolution sequencing of telomere fusion events, this study has identified these known TVRs in addition to other less abundant repeat motifs.

The mechanism for generation of TVRs remains unclear. Some studies propose that they are product of the telomere repeat misincorporation by telomerase (Lee *et al.*, 2014a). In this situation, a random proportion and distribution of TVRs would be expected for each telomere. In contrast, a unique composition of TVRs has been identified for XpYp, 12q, 16p, 16q and 17p telomeres, where these TVRs were considered to be propagated via intra-

allelic mechanisms including replication slippage or unequal sister chromatid exchange (Baird, 1995, Coleman *et al.*, 1999, Baird *et al.*, 2003, Letsolo *et al.*, 2010). Consistently, in this study, a greater proportion of variant repeats was identified in 5p compared with XpYp and this differential provides an explanation for the differences in the relative TL distributions observed. These results also suggest that differences in TLs within the same cell may, at least in part, be explained by the unique diversity of TVRs at distinct chromosome ends. Thus, it is likely that TVRs arise from replication errors and that only those restricted to the proximal telomeric sequence are maintained. Moreover, if TVRs propagate through sister chromatid exchange, it is likely that they are maintained at the specific chromosome ends and therefore that individual telomeres have a distinct TVR-pattern associated.

A potential impact of the extent of TVRs may be related to chromatin modifications modulated by TL. Telomere position effect over long distance (TPE-OLD) (Robin *et al.*, 2014) has been proposed to epigenetically silence telomerase through a telomeric loop when telomeres are long (Kim *et al.*, 2016). However, when telomeres become short, TPE-OLD mechanism could be implicated in reactivation of telomerase in cancer. It is interesting to speculate that the 5p telomere may contain more TVR, that make this telomere significantly longer than other telomeres, to help repress telomerase in somatic cells. However, this would depend on the affinity of the shelterin components to bind the different types of TVRs and the specific TVR interspersions patterns.

6.2.2 Telomere dysfunction and fusion, including the 5p telomere, in CLL patient samples and in HCT116 human colorectal cancer cell line

Telomere shortening is considered to function as a tumour suppressor mechanism through the induction of senescence and restriction of the replicative lifespan of normal somatic cells (Harley *et al.*, 1990, Bodnar *et al.*, 1998). However, in the absence of functional DNA damage cell-cycle checkpoints, dysfunctional telomeres can initiate cycles of breakage-fusion-bridge (BFB) that could result in large-scale genomic rearrangements that drive cancer progression (McClintock, 1941, Ma *et al.*, 1993, Maciejowski and de Lange, 2017). For malignant progression, TL stabilisation is required, which occurs by telomerase reactivation in 85% of human cancers (Kim *et al.*, 1994, Meyerson *et al.*, 1997). Different mechanisms of telomerase reactivation include point mutations in the *hTERT* promotor, gene amplifications or translocations that result in *hTERT* being proximal to an enhancer that can drive gene expression (Horn *et al.*, 2013, Huang *et al.*, 2013, Bell *et al.*, 2015,

Valentijn *et al.*, 2015). In CLL, genomic rearrangements of the 5p chromosome arm including *hTERT* have been associated with telomerase upregulation (Nagel *et al.*, 2010, Schilling *et al.*, 2013, Salaverria *et al.*, 2015). Since the *hTERT* locus is proximal to the 5p telomere (1.2Mb away), it is plausible that 5p-5p sister chromatid telomere fusions could initiate *hTERT* amplification. This study investigated telomere dysfunction and fusion at the 5p telomere to determine whether this mechanism can be implicated in CLL progression. Telomere fusions have previously been detected in CLL patients with short telomeres in association with increased genomic instability (Lin *et al.*, 2010). The adaptation of the telomere fusion assay (Capper *et al.*, 2007) to include the 5p chromosome end enabled, for the first time, the detection of 5p telomere fusions in CLL patient samples. Thus, 172 fusion events involving the 5p telomere have been identified, including inter-chromosomal fusions with other telomeres and non-telomeric loci, which could potentially translocate *hTERT*. Furthermore, 5p-5p intra-chromosomal telomere fusions with the potential to drive gene amplification were identified. Although it is unknown whether these ongoing events can provide further selective advantage in pre-established telomerase-expressing CLL cells, they provide evidence for a biological process that could occur in pre-malignant cells during a telomere-driven crisis or contribute to treatment-associated relapse.

The frequency of 5p fusions detected was likely to be under-represented. This is because within the 5p telomere-adjacent DNA, a CpG island (~0.9 Kb) was identified which prevented the amplification and sequencing of telomere fusions that included this region. GC-rich sequence can form secondary structures and therefore are usually refractory to long-range PCR (McDowell *et al.*, 1998).

Amplification of the 5p chromosome end including the *hTERT* locus and upregulation of telomerase activity was observed in an HCT116^{DN-*hTERT*} cell model that escaped from a telomere-driven crisis, (Jones *et al.*, 2014). Telomere fusions including the 5p telomere were detected during the crisis state prior to 5p amplification; however, there was no evidence of 5p duplication or LOH from the TL profile after crisis. Since the 5p TL distribution was very heterogeneous post-crisis, which may be indicative of distinct subclonal populations, it is conceivable that telomerase was not reactivated throughout the population.

6.2.3 Telomere fusion frequency and prognosis in CLL patients with short telomeres

Given the diverse clinical outcome of patients with CLL, it is essential to establish accurate predictors of patient survival. It has long been considered that TL could be a prognostic tool in CLL (Bechter *et al.*, 1998) and several studies have attempted to combine TL with distinct prognostic markers including Binet stage, *IGHV* gene status and cytogenetics (Grabowski *et al.*, 2005, Rossi *et al.*, 2009). The development of STELA allowed the measurement of single telomeres to precisely determine TL distributions with high resolution (Baird *et al.*, 2003). Using STELA in CLL patient samples it was observed that shorter TL distributions, within the range at which telomere fusions can be detected, accurately identified patients with poorer prognosis (Lin *et al.*, 2010, Lin *et al.*, 2014). Short telomeres have also been linked to elevated genome instability and potential for Richter transformation (Roos *et al.*, 2008, Lin *et al.*, 2010). TL predicted response to treatment in CLL patients from the ARCTIC and ADMIRE cohorts (Norris *et al.*, manuscript in preparation). However, the impact of the frequency of fusions on patient survival had not been reported. Thus, in this study it was investigated whether the presence or the frequency of fusions could resolve additional subdivisions within the subgroup of CLL patients with short TL. The *IGHV* gene mutation status, the widely used prognostic marker, failed to further stratify patients with short telomeres for progression-free survival (PFS) and overall survival (OS). Whilst not statistically significant, 15 month shorter PFS and reduced OS were observed within a subgroup of patients with the highest frequency of fusions ($>4.20 \times 10^{-5}$) than the subgroup without fusions. Therefore, further research with a larger sample size or a more sensitive technique is required to clarify whether the frequency of fusions provides valuable prognostic information, in parallel with TL, in CLL patients.

The presence of deletion of 11q (*ATM*) and 17p (*TP53*) are valuable prognostic markers in CLL (Krober *et al.*, 2002). Inactivation of *TP53* by mutation or 17p deletion in CLL patients has been associated with short telomeres, high telomerase expression, chromosome end-to-end fusions and high genomic instability (Guieze *et al.*, 2016). Surprisingly, no association between deletion/mutation status and fusion frequency was identified in this study. It remains possible that not all telomere fusions were captured or that they occurred earlier in the progression of CLL and are absent in the expanded clones. In addition, the proportion of CLL patients screened for *TP53* and *ATM* mutations was too small to draw any significant conclusions: 2/30 and 3/13 patients with mutated genes, respectively.

6.2.4 The impact of telomere dysfunction and fusion on the CLL genome

It is known that dysfunctional telomeres can undergo BFB cycles and initiate genomic instability including amplifications, deletions, translocations, chromothripsis and tetraploidy (Capper *et al.*, 2007, Maciejowski and de Lange, 2017), most of which are commonly found in the CLL genome (Ghamlouch *et al.*, 2017). As discussed in the previous section, CLL patients with short and fusogenic telomeres associate with increased genomic instability and worse prognosis. Altogether, this suggests that telomere dysfunction may play an important role in shaping the CLL cancer genome. Although Sanger sequencing of reamplified telomere fusions has previously been reported (Lin *et al.*, 2010), a thorough characterisation of such events had not been performed. This study has allowed the high-resolution characterisation of 914 fusion events from 9 CLL patients, revealing distinct signatures of telomere fusions.

This study revealed 7 complex telomere fusions that incorporated up to 4 distant loci, indicating genome-wide instability. Complex telomere fusion events had previously been observed in MRC5 fibroblasts and the HCT116 colorectal cancer cell line (Capper *et al.*, 2007, Letsolo *et al.*, 2010, Liddiard *et al.*, 2016). In agreement with Liddiard *et al.*, this study revealed that telomere fusions can incorporate genomic loci, predominantly with coding sequence. Surprisingly, CLL-related genes and other oncogenes, as well as mitochondrial DNA were also incorporated into telomere fusions. Boulianne *et al.*, observed in B cell precursors that increased DNA damage and fragility, characterised by DSBs, concentrated at leukaemia specific genes during malignant progression because of an elevated transcription of these loci (Boulianne *et al.*, 2017). Transcription-associated genomic fragility was related to exposed ssDNA, associated with R-loop formation and induction of DSB mediated by Topoisomerase 2 during rapid transcriptional activation (Boulianne *et al.*, 2017). It is therefore credible that genes actively transcribed during the development of CLL are more prone to damage and hence may become incorporated into telomere fusions that result from the inappropriate repair of DSBs. Telomere fusion events could be detrimental for cell viability since they may disrupt genes required for cell survival or contribute to the cellular burden of genomic instability that is poorly-tolerated. However, fusions may also provide a selective advantage for malignant transformation and disease progression. They could potentially disrupt tumour suppressor genes, recombine distant locations and reactivate oncogenes, including *hTERT*, shaping the CLL genome.

6.2.5 Telomere dysfunction and fusion potentially drive clonal evolution and intra-tumour heterogeneity

The cancer genome is subjected to mutation and large-scale genomic alterations that can drive tumour evolution with the selection of the fittest clones over time (Nowell, 1976, Greaves and Maley, 2012). In CLL, elevated inter- and intra-patient tumour heterogeneity has been observed (Crossen *et al.*, 1993, Landau *et al.*, 2013, Landau *et al.*, 2014). The evolution and outcome of the patient's disease can be influenced by the co-existence of multiple subclones within the same patient, which is associated with adverse clinical outcome (Crossen *et al.*, 1993, Landau *et al.*, 2013, Landau *et al.*, 2014). In addition, therapy and tumour microenvironment (including hypoxia, acidosis, ecological niche and immune system) can accelerate this process selecting for resistant subclones and resulting in patient relapse (Puente and Lopez-Otin, 2013, Ojha *et al.*, 2015, Del Giudice *et al.*, 2016, McGranahan and Swanton, 2017).

In this study, an increased proportion of CLL patients within the subgroup with the highest frequency of fusions exhibited bimodal TL distributions compared with the ARCTIC and ADMIRE cohort (67% vs. 4%). The presence of distinct subclonal populations with different TLs was confirmed for one of the CLL patients. In this case, it is conceivable that the subclone with shorter telomeres evolved from the primary clone, with subsequent telomere erosion over time as the subclonal population expanded. It remains possible that the bimodal TL distribution observed in CLL patients with the highest frequency of fusions, in addition to heterogeneous TL distributions with large standard deviation, reflect multiclonal populations. Further evidence supporting telomere dysfunction and fusion driving tumour heterogeneity is provided by the comparison of the 5p TL distribution in an HCT116^{DN-hTERT} clone prior to and after escape from a telomere-driven crisis. A clonal population with a very homogeneous TL distribution resulted in an extremely heterogeneous with potentially distinct TL clusters indicative of multiple subclones.

These results provide insights into the impact of telomere dysfunction and fusion as a mechanism for the production of genetically heterogeneous cells that may drive multiclonal evolution (intra-tumour heterogeneity) and a more aggressive disease.

6.2.6 Excessive genetic instability and decreased tolerance to DNA damage may negatively impact on cancer cell survival, improving patient outcome

Chromosomal instability (CIN) is generally associated with poor prognosis in human cancers since it can provide some survival advantage to the cell. However, severe CIN has been shown to have a negative impact on cancer cell viability as it may induce cell death (Cahill *et al.*, 1999). Extreme levels of CIN have been associated with better patient prognosis, compared to intermediate CIN, in breast, ovarian, gastric, and non-small cell lung cancer (Birkbak *et al.*, 2011). Similarly, the number of distinct subclones within a tumour may also influence patient survival. An increased mortality risk was observed in patients that presented 2 to 4 distinct clones; however, mortality risk decreased when more than 4 subclones coexisted (Andor *et al.*, 2016).

In this investigation, a case study of a 79 year old CLL patient with elevated genomic instability but an indolent disease was presented. Analysis of the patient's CLL-B cells revealed the highest frequency of telomere fusions from a cohort of 276 CLL patients, intra-tumour heterogeneity and a CLL genome characterised by chromosomal rearrangements and CLL driver mutations.

Mutations in CLL drivers and in genes involved in the oncogenic MAPK/ERK signalling pathway (upregulated in a subset of CLL patients (Ferreira *et al.*, 2014)) were identified in this patient, suggesting that DNA damage occurs in genes that are actively transcribed (Boulianne *et al.*, 2017). The common CLL 6q21 and 13q14 deletions were identified, but found to arise from a previously unreported 6q:13q translocation. Interestingly, a 17p copy neutral LOH or UPD with a homozygous *TP53* mutation associated with high-risk CLL was revealed. Since 17p telomere fusions were detected, it is conceivable that dysfunction at this telomere initiated the 17p UPD. Altogether, the genomic rearrangement and high-impact mutations described were clonal and not exclusive to the subclonal population. However, it remains possible that some subclonal events were not detected.

A mutation in *REV3L* was identified 2.2Mb downstream (towards the telomere) of the 6q21 deletion; however, it was not possible to determine whether it was located on the translocated chromatid since the deletion was mono-allelic and the mutation heterozygous. This novel mutation in *REV3L* is predicted to result in a truncation in the catalytic subunit of polymerase ζ that could compromise translesion synthesis and

contribute to the mutational burden of the cell. This *REV3L* mutation could, therefore, sensitise cells to DNA damage leading to increased genomic instability that results in cell death despite mutated *TP53*. Previous research has shown that *REV3L* loss-of-function sensitises cells to chemotherapeutic agents including cisplatin and is associated with increased DSBs and anaphase bridges, whilst its overexpression confers resistance to DNA damaging agents (Bhat *et al.*, 2013, Lee *et al.*, 2014b, Yang *et al.*, 2015, Lange *et al.*, 2016, Huang *et al.*, 2016).

The novel *REV3L* mutation identified in this CLL patient could hence be related to the increased frequency of fusions and, together with the increased genomic instability and the intra-tumour heterogeneity, may explain the patient's indolent disease resulting from increased cancer cell death. Further research is required since *REV3L* may be a promising therapeutic target to alter chemosensitivity in several cancers, including CLL.

6.2.7 Potential therapeutic interventions to prevent cells from escaping crisis

Targeting the DNA repair mechanisms that mediate telomere fusions may provide a potential cancer therapeutic strategy that could limit clonal evolution.

Telomere fusion events lacking telomere repeats have previously been detected in cells undergoing crisis in culture, as well as in CLL and breast cancer samples (Capper *et al.*, 2007, Lin *et al.*, 2010, Letsolo *et al.*, 2010, Roger *et al.*, 2013). In addition, subtelomeric asymmetrical deletion has been observed at fused sister chromatids (Liddiard *et al.*, 2016). Consistent with previous research, the analysis of 5p, 17p and XpYp intra-chromosomal fusions in CLL patients suggests that stochastic DSB and/or differential resection can occur at subtelomeric DNA before ligation. In addition, Liddiard *et al.*, proposed C-NHEJ and A-NHEJ as DNA repair mechanisms primarily mediating inter- and intra-chromosomal telomere fusions, respectively. In CLL, microhomology at the junction of telomere fusions consistent with A-NHEJ has previously been observed (Lin *et al.*, 2010). In this study, the characterisation of 739 telomere fusion junctions revealed a differential usage of microhomology for the distinct types of telomere fusion events detected, implicating divergent DNA repair mechanisms. C-NHEJ most likely mediates head-to-head fusions containing telomeric sequence since microhomology usage was low or undetected at the fusion junctions. In contrast, A-NHEJ probably facilitates sister chromatid fusions and inter-chromosomal fusions with genomic loci since microhomology usage was evident at fusion junctions. An additional DNA repair mechanism, potentially SSA that depends on higher

usage of microhomology, may be responsible for a subgroup of telomere fusions with genomic loci.

Jones *et al.*, showed that LIG3 was required to escape a telomere-driven crisis, mediated by A-NHEJ and cells that escaped exhibited increased proportions of intra-chromosomal compared with inter-chromosomal telomere fusions (Jones *et al.*, 2014). Targeting components of the A-NHEJ pathway, including LIG3, LIG1 and PARP1, have been shown as promising therapeutic targets in neuroblastoma (Newman *et al.*, 2015). Currently, PARP inhibitors are effective in *BRCA1* and *BRCA2*-deficient breast and ovarian cancer that are compromised for homologous recombination (HR) repair (Livraghi and Garber, 2015, Meehan and Chen, 2016). PARP inhibitors block ss-DNA repair mechanism at the site of damage that results in increased DSBs. The repair is then mediated by error-prone NHEJ that can cause genomic instability at a level that is incompatible with cell viability (Lord and Ashworth, 2016). Another strategy investigated for the treatment of leukaemia has been the inhibition of cell cycle checkpoint components (ATM/ATR/CHK1/CHK2/WEE1) that have a primary role in the DDR. These inhibitors are mostly used to enhance the effectiveness of standard chemotherapy by blocking the response of tumour cells to genotoxic agents and result in accumulation of DNA damage, arrest and cell death (Ghelli Luserna di Rora *et al.*, 2017). The long-term effects on normal healthy cell populations remain to be determined, in particular to exclude the possibility that they could undergo malignant transformation (Ghelli Luserna di Rora *et al.*, 2017).

The benefit of patients with *BRCA*-mutated cancers treated with PARP inhibitors, highlights how elevating genomic instability to lethal levels in genetically unstable cancer cells can be exploited as a therapeutic target (Lord and Ashworth, 2016). In CLL, a recent report identified that *ATM*-deficient mice were sensitive to PARP inhibition (Knittel *et al.*, 2017). Further research should be undertaken to study the impact of PARP inhibitors in CLL-B cells. PARP inhibitors may be effective on their own in a subset of patients with deletion of 11q or *ATM* mutation, potentially compromised for HR, since ATM recruits BRCA1 to DSBs and promotes HR (Isono *et al.*, 2017). In cells with no previous evidence for HR-deficiency, PARP inhibitors could be used in combination to molecules that interfere with HR, like RAD51 inhibitors (Wang *et al.*, 2012).

Since LIG3-mediated A-NHEJ was required to escape a telomere-driven crisis (Jones *et al.*, 2014) and most telomere fusions in CLL patients are potentially mediated by A-NHEJ, it will be important to investigate whether PARP inhibitors may also be effective in those patients

with the highest frequency of fusions. Otherwise, they may need to be used in combination with conventional DNA damaging agents. *REV3L* loss-of-function was shown to compromise translesion synthesis and HR (Huang *et al.*, 2016), thus clinical targeting of *REV3L* could be explored to enhance cancer cell sensitivity to PARP inhibitors. Patients with tumours that exhibit short dysfunctional telomeres may represent good candidates for these novel agents or treatment combinations.

6.3 CONCLUSIONS AND FUTURE DIRECTIONS

- In this project, tools to successfully study the 5p telomere that is proximal to the *hTERT* locus have been provided: 5p STELA, TVR-PCR and the telomere fusion assay at 5p. 5p STELA allows the determination of TL profiles and LOH at the 5p telomere. Furthermore, the adapted telomere fusion assay is an improvement on the existing technique that allows detecting increased number of fusions including the 5p chromosome arm. Together, both assays may turn out to be an informative diagnostic tool for measuring TL and genomic instability in patient samples.
- The 5p TL distributions are similar to the XpYp and 17p telomeres. Distinct telomeres may contain the same amount of canonical repeats but different proportions of TVRs. It has been proposed that TPE-OLD at the 5p telomere may repress telomerase expression (Kim *et al.*, 2016). Research questions that should be asked include whether the 5p telomere contains greater proportion of TVRs or specific TVR interspersions that are capable of binding TRF1 and TRF2, resulting in a longer telomere to further protect somatic cells from telomerase reactivation. A comparison of the TL and sequence composition of distinct telomeres amplified with STELA should be undertaken in an increased cohort size to precisely quantify canonical and variant repeats.
- Telomere dysfunction and fusion at the 5p telomere has been identified, for the first time, in CLL patient samples. Sequencing 5p telomere fusions has revealed that this chromosome end can fuse to other telomeres and genomic loci, which could potentially translocate the *hTERT* locus near an enhancer. Most importantly, intra-chromosomal fusion events that may provide a potential mechanism of *hTERT* amplification were identified. Further research should be undertaken to establish whether telomere dysfunction and fusion at 5p precedes *hTERT* reactivation and

facilitating the escape from crisis. A comparison of the ability of HCT116 DN-*hTERT* cells, with or without targeted 5p DSBs induced using TALEN or CRISPR/Cas9 technology to escape crisis and reactivate *hTERT* could be performed. 5p fusion primers could be designed upstream of the 5p TALEN or CRISPR/Cas9 breakpoint to capture and sequence telomere fusions. This would allow investigation of whether instability at this telomere leads to *hTERT* rearrangements that drive malignant progression.

- 17p telomere fusions were detected in the CLL patient with the 17p CN-LOH or UPD with a homozygous *TP53* mutation. It should be further investigated whether telomere dysfunction can result in this genomic rearrangement. A potential strategy could be to experimentally induce DSBs using the 17p telomere-specific TALEN system in HCT116^{DN-*hTERT*} cells and investigate 17p LOH and UPD in clones escaping replicative senescence most likely from the loss of function of the tumour suppressor *TP53*.
- Varied frequencies of telomere fusions were identified amongst CLL patient samples with short telomeres. Further investigation would allow to clarify whether the fusion frequency differences arise from specific mutations in cell cycle checkpoints or dysfunctional sheltering components in each particular patient. A panel of genes, including *TP53*, *REV3L*, *POT1*, *TRF2*, *ATM* and *ATR*, should be sequenced in the cohort of CLL patients.
- The frequency of telomere fusions failed to provide further prognostic information among a subgroup of CLL patients with short telomeres. However, since shorter PFS and reduced OS were observed within a subgroup of patients with the highest frequency of fusions ($>4.20 \times 10^{-5}$) than the subgroup without fusions, further research with an increased cohort size should be undertaken. Moreover, it might be useful to develop a more sensitive technique to detect telomere fusions.
- The frequency of bimodal TL distributions was enriched in a subset of patients with the highest frequency of fusions, potentially reflecting intra-tumour heterogeneity. This was confirmed for one of the CLL patients but to validate this hypothesis, the presence of distinct subclones should be investigated in the remaining CLL patients

with bimodal TL distribution. In addition, it remains possible that patients with heterogeneous TL profiles with greater standard deviations also represent the tumour heterogeneity within the sample. Together with the previous finding, it is suggested that increased telomere dysfunction and BFB cycles provide the genomic diversity required to drive intra-tumour heterogeneity which results in a more aggressive disease. This would provide additional prognostic value to the measurement of TL and the frequency of fusions; however, further studies need to be undertaken to validate this hypothesis.

- An exception to the previous hypothesis has been observed in a case study. Thus, it is interesting to propose that dysfunctional *REV3L* may potentially be implicated in the increased frequency of fusions and sensitivity to DNA damage in the patient CLL-B cells that may have resulted in the patient's indolent disease. To investigate the impact of this truncated protein on Pol ζ function, a MEC1 CLL cell line (Stacchini *et al.*, 1999) model with WT and mutated *REV3L* (Gln1426*) could be established using TALEN or CRISPR/Cas9 technology. The tolerance to translesion synthesis, the impact on telomere fusions and cellular fitness would be investigated. In addition, WT and *REV3L*-mutant cells could be treated with cisplatin, other DNA damaging agents or cell cycle checkpoint inhibitors to investigate whether this protein may provide a potential therapeutic target in CLL. In the future, those CLL patients with the highest frequency of fusions might benefit from Pol ζ functional inhibition to elevate the levels of genomic instability beyond the threshold tolerated by the cancer cell. In combination with other DNA damaging agents, this approach may also be effective in other cancer patients.
- In addition, further research should be taken to investigate the impact of PARP inhibitors in patients with dysfunctional telomeres, particularly those ATM-deficient CLL patients that may be compromised for HR repair. In the future, screening patients with 11q deletion or mutated *ATM* could allow selecting for more tailored treatments.
- Telomere dysfunction and fusion, originating from extremely short or unprotected telomeres, likely facilitates the acquisition of the rearrangements found in the CLL genome that lead to malignant progression. The high-resolution sequencing of CLL

patient telomere fusion events has revealed genomic loci that can be incorporated into telomere recombinations. Such events could be expected to be detrimental to normal cell gene expression, DNA replication and cell cycle progression. However, certain molecular recombinations have the potential to confer a selective advantage if they disrupt growth suppressor genes or reactivate genes that facilitate immortalisation, including *hTERT*. To determine the impact and functional consequences of telomere-genomic fusions in CLL, specific areas of the genome identified as susceptible to recombination with telomeres in CLL patient samples could be destabilised. This could be achieved by using TALEN or CRISPR/Cas9 technology in the MEC1 CLL cell line (Stacchini *et al.*, 1999). Subsequent to targeted genomic disruption, telomere fusion frequency, cell viability, cell cycle progression and RNA or protein expression should be studied.

- Altogether, telomere dysfunction and fusion, originating from extremely short or unprotected telomeres, likely facilitates the acquisition of the rearrangements found in the CLL genome that lead to malignant progression.

6.3.1 Final conclusions

The final conclusions obtained from this Ph.D. project are summarised in **Figure 6.1**. Telomere dysfunction and BFB cycles can initiate genome-wide instability as evidenced from the high-throughput sequencing of 914 telomere fusion events from 9 CLL patient samples. This has the potential to drive large-scale genomic rearrangements including deletions, amplifications and translocations that are commonly detected in CLL patients' genome.

Telomere fusions provide a source for genetic variability that can result in two distinct cellular outcomes. Fusions can deliver a selective advantage to the cell for malignant progression, driving clonal evolution or intra-tumour heterogeneity. Evidence to support this is provided by the increased frequency of CLL patients with bimodal TL detected amongst the subgroup with the highest frequency of fusions compared with the total ARCTIC and ADMIRE cohort (67% vs. 4%). Biclinality or multiclinality was validated for one of the CLL patients but further research should be taken to investigate multiclinality within the other CLL patient samples. In addition, an increased frequency of fusions and intra-tumour heterogeneity may result in a more aggressive disease.

An alternative outcome from a telomere-driven crisis is cell death, resulting from the increased and unsustainable genomic instability. Thus, I propose that this may be disadvantageous to the cancer cell and could be exploited for therapeutic benefit by targeting DNA repair mechanisms in cells already sensitised to DNA damage. In this study, the CLL patient with increased genomic instability, high fusion frequency and an indolent disease was found to possess a novel *REV3L* gene mutation predicted to result in a loss-of-function of Pol ζ , thereby potentially sensitising the CLL cells to DNA damage. The functional consequences of this mutation should be further investigated since it may prove a useful therapeutic target in CLL.

Altogether, in this Ph.D. project I have illustrated how telomere dysfunction can potentially shape the CLL genome and modify the course of the disease, as well as identified novel therapeutic targets that may prove useful in subsets of CLL patients.

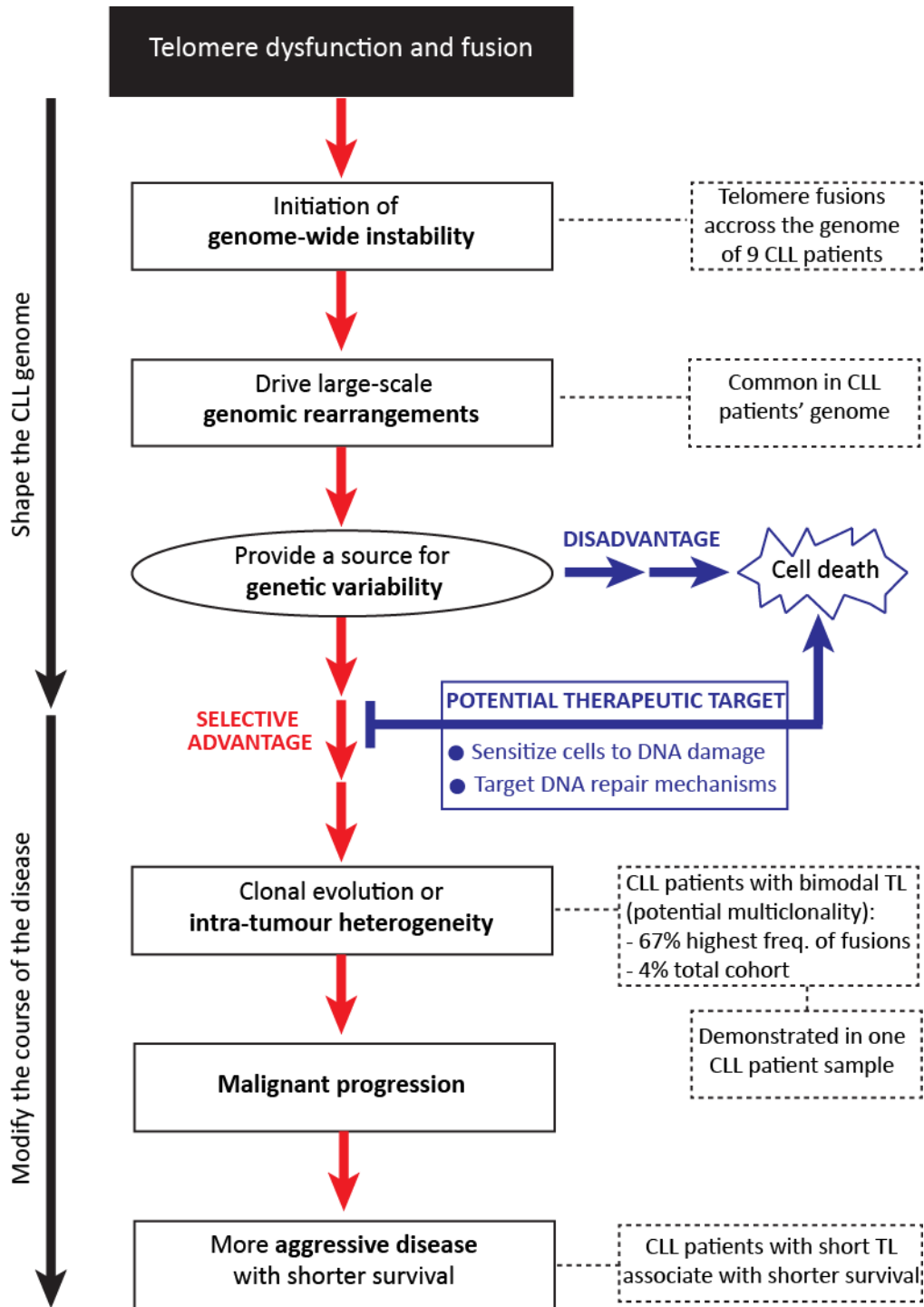
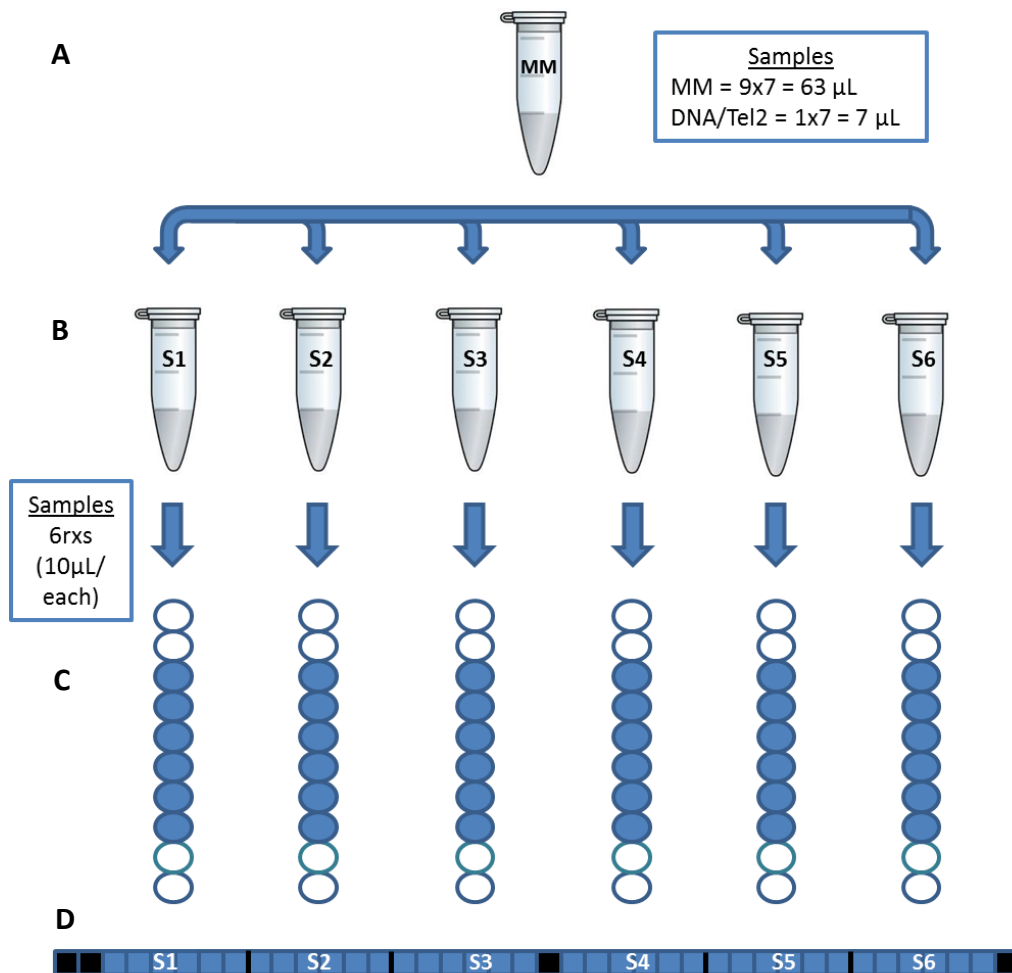


Figure 6.1. The impact of telomere dysfunction and fusion on Chronic Lymphocytic Leukaemia.

Summary of the findings identified from this Ph.D. project, providing evidence on how dysfunctional telomeres can shape the CLL genome and modify the course of the disease.

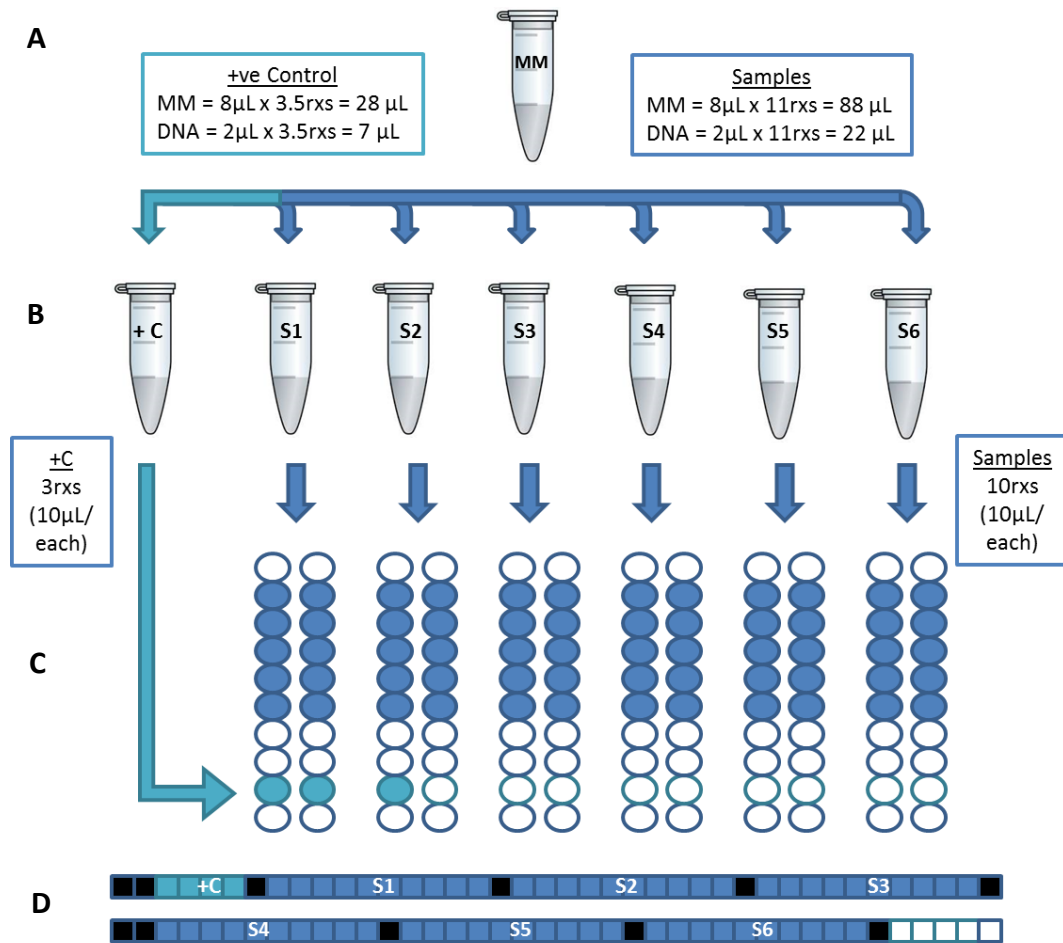
APPENDIX

SUPPLEMENTARY FIGURES



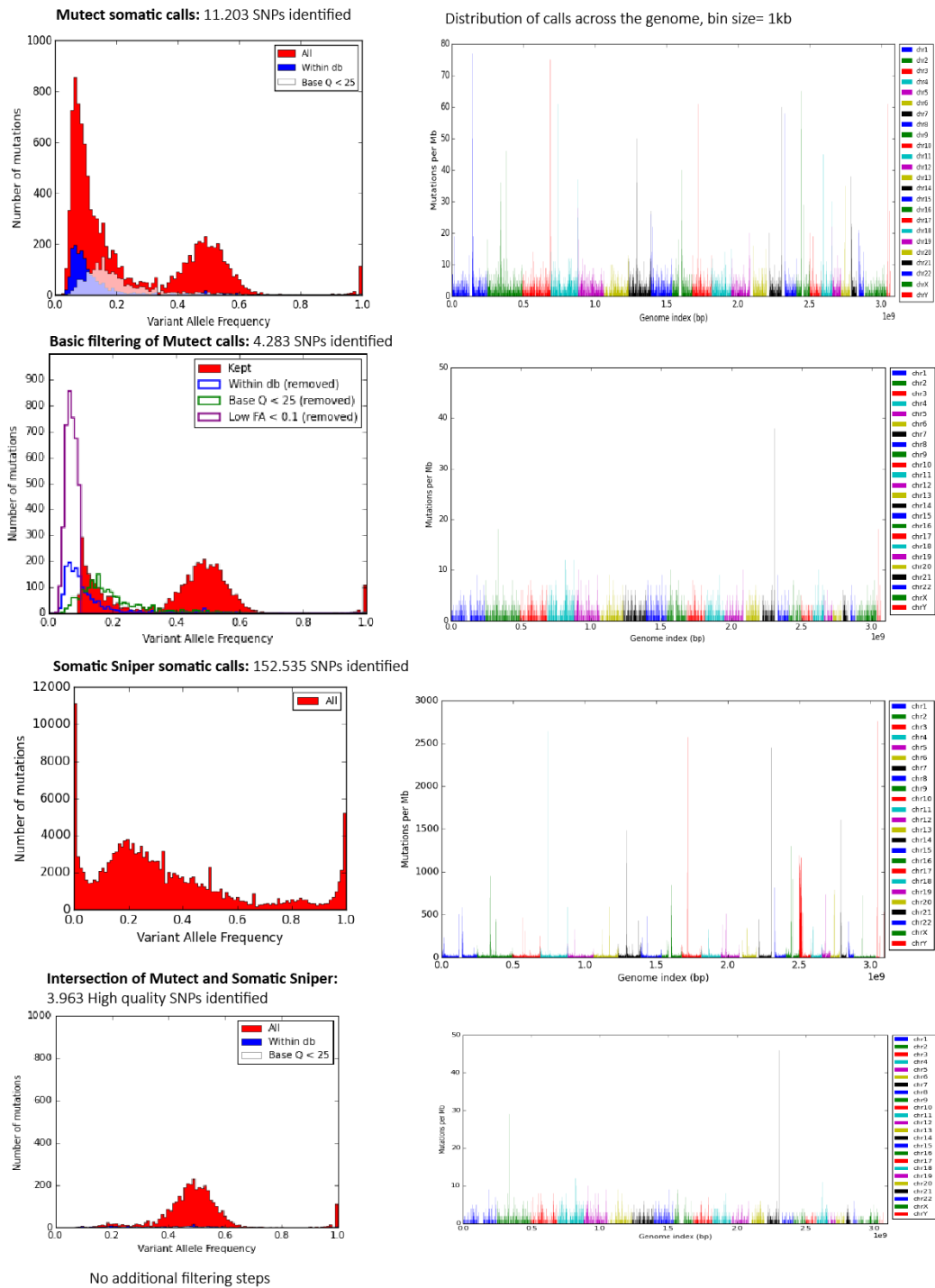
Supplementary Figure 1. STELA protocol for 6 samples.

(A) Master mix (MM) calculated for 6 + 1 samples to allow for additional volume. **(B)** 6 reactions (rxs) were prepared for each sample; 63µL of MM was transferred into a tube containing 7µL of the DNA/Tel2 mix. **(C)** 10µL of each reaction was loaded into the 8-well PCR strips that were transferred into the 96-well thermo cycler for STELA PCR. **(D)** The 6 STELA PCR rxs prepared per each of the 6 samples (blue squares) together with 4 wells used for the DNA ladders (black square) that were loaded into a 40-well agarose gel. Usually, several STELA agarose gels were run at the same time. In such occasions the preceding process was escalated for the TL analysis of multiples of 6 numbers of samples.



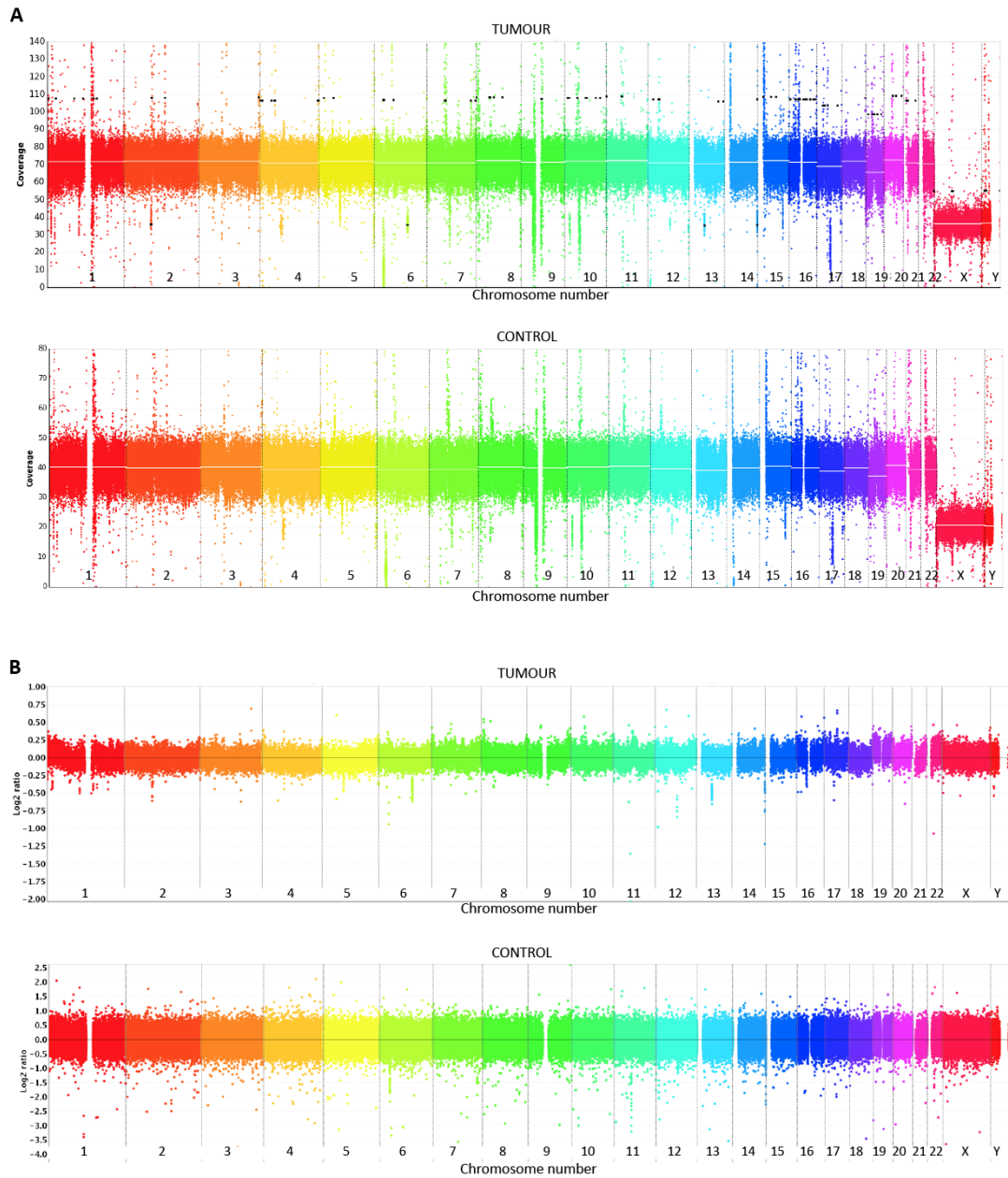
Supplementary Figure 2. Telomere fusion protocol for 6 samples.

(A) Master mix (MM) calculated for the positive control, 6 samples and additional volume. (B) For the positive control, 3.5 reactions (rxs) were prepared so 28 μL of MM was transferred into a tube containing 7 μL DNA. For the 6 samples, 11rxs were prepared for each sample so 88 μL of MM was transferred into a tube containing 22 μL DNA. (C) 10 μL from 3 rxs for the positive control, and 10 μL from 10 rxs of each of the 6 samples were loaded into the 8-well PCR strips. The telomere fusion PCR rxs were transferred into the 96-well thermo cycler. (D) The 10 PCR rxs prepared for each of the 6 samples (blue squares) together with the 3 rxs of the positive control (light blue squares) were loaded into a 2x40-well agarose gel, alongside the DNA ladders (black square). The rows were separated by 20cm in the 40cm-long gel. The preceding process was escalated for the telomere fusion analysis of multiples of 6 numbers of samples as several agarose gels were run at the same time.

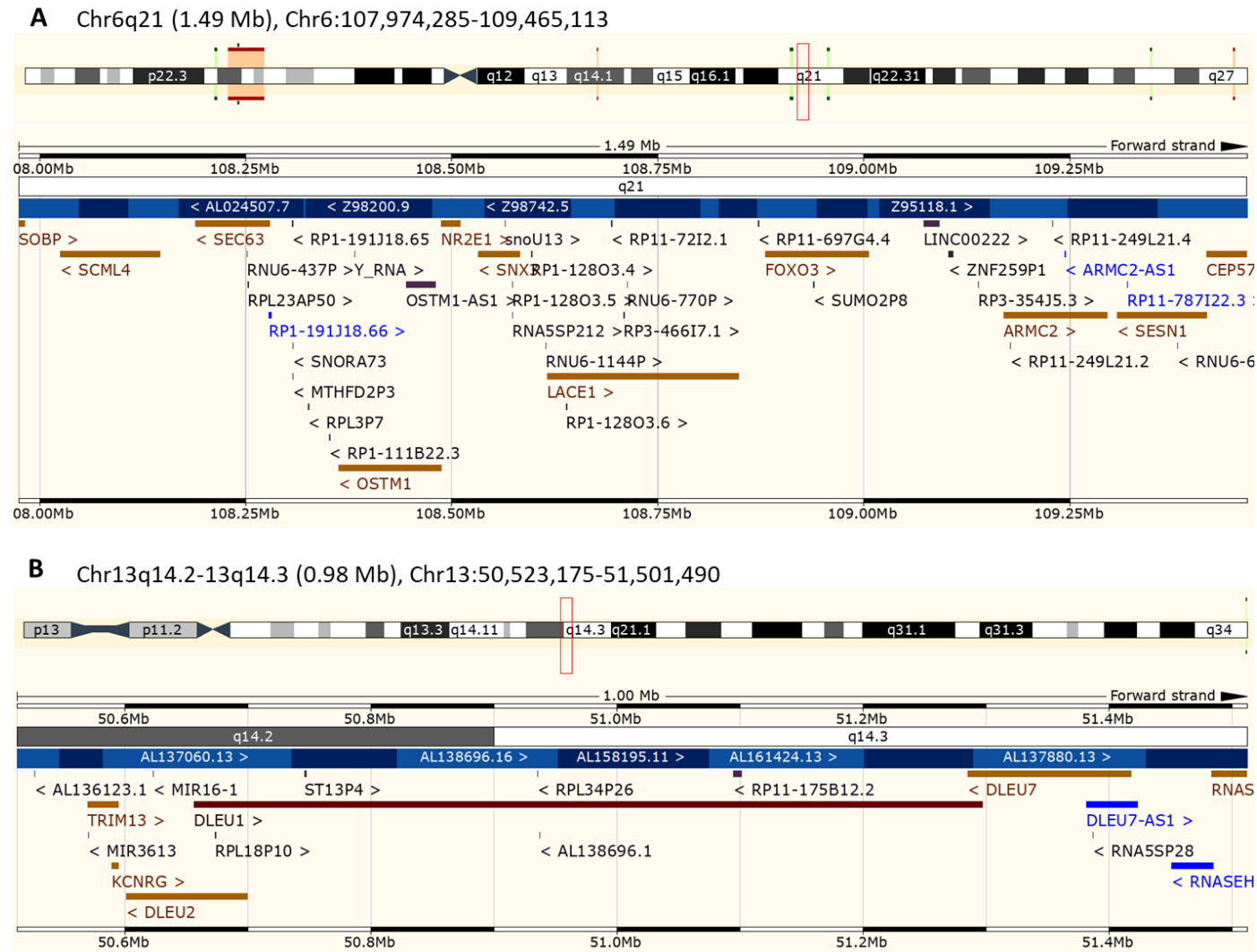


Supplementary Figure 3. Identifying SNVs present on the tumour sample.

For increased accuracy a combination of Mutect and Somatic Sniper was used to detect SNVs from DB17 patient tumour/control samples. SNVs from the intersection were selected. Detected in patient DB17 CLL-B cells.

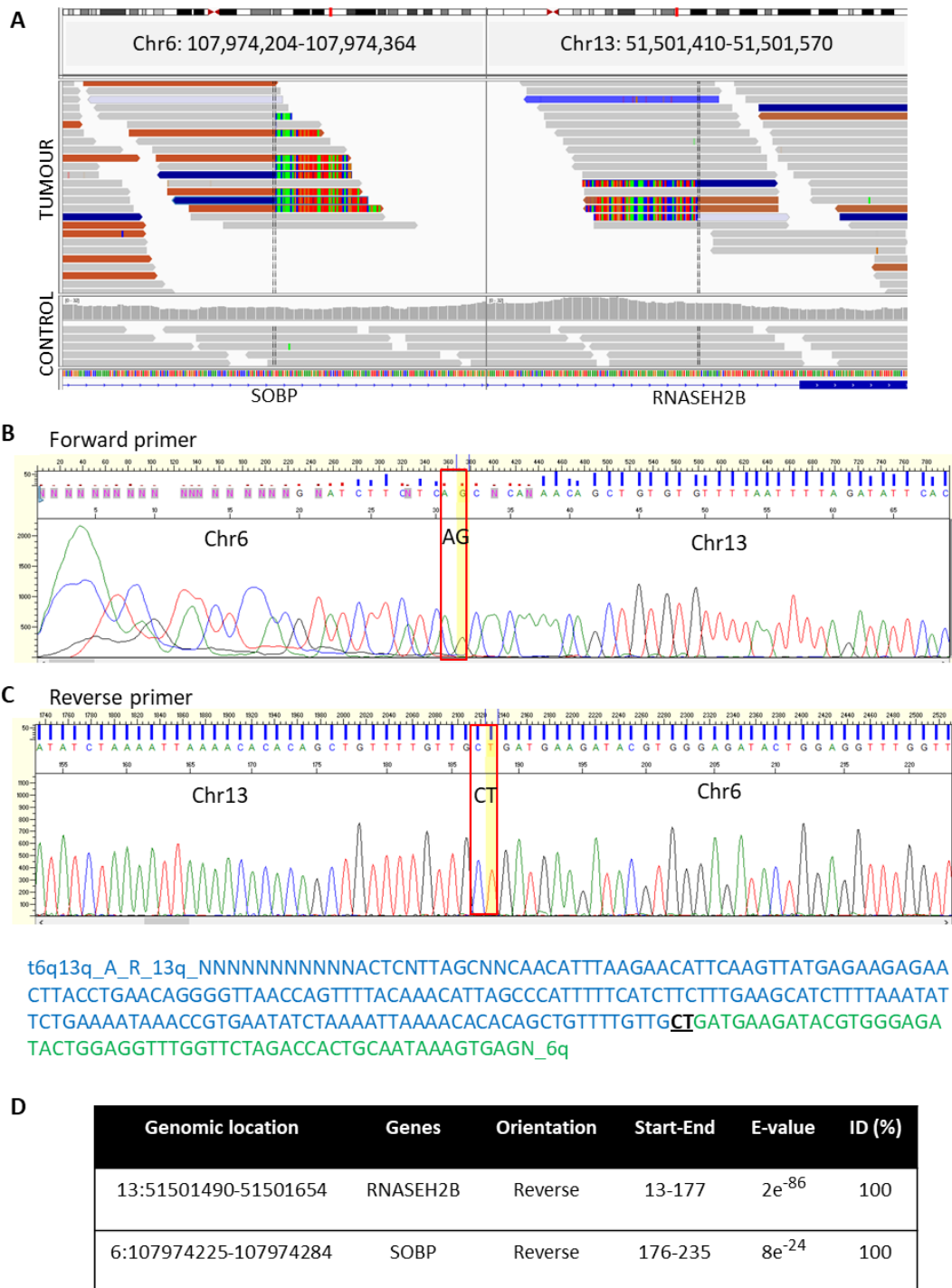


Supplementary Figure 4. Copy Number Variation identified across the genome of DB17 CLL-B cells. CNV detected by **(A)** cn.MOPS and **(B)** array CGH for the DB17 tumour and control samples. CNV study was performed in **(A)** 2014 and **(B)** 2010. Detected in patient DB17 CLL-B cells.



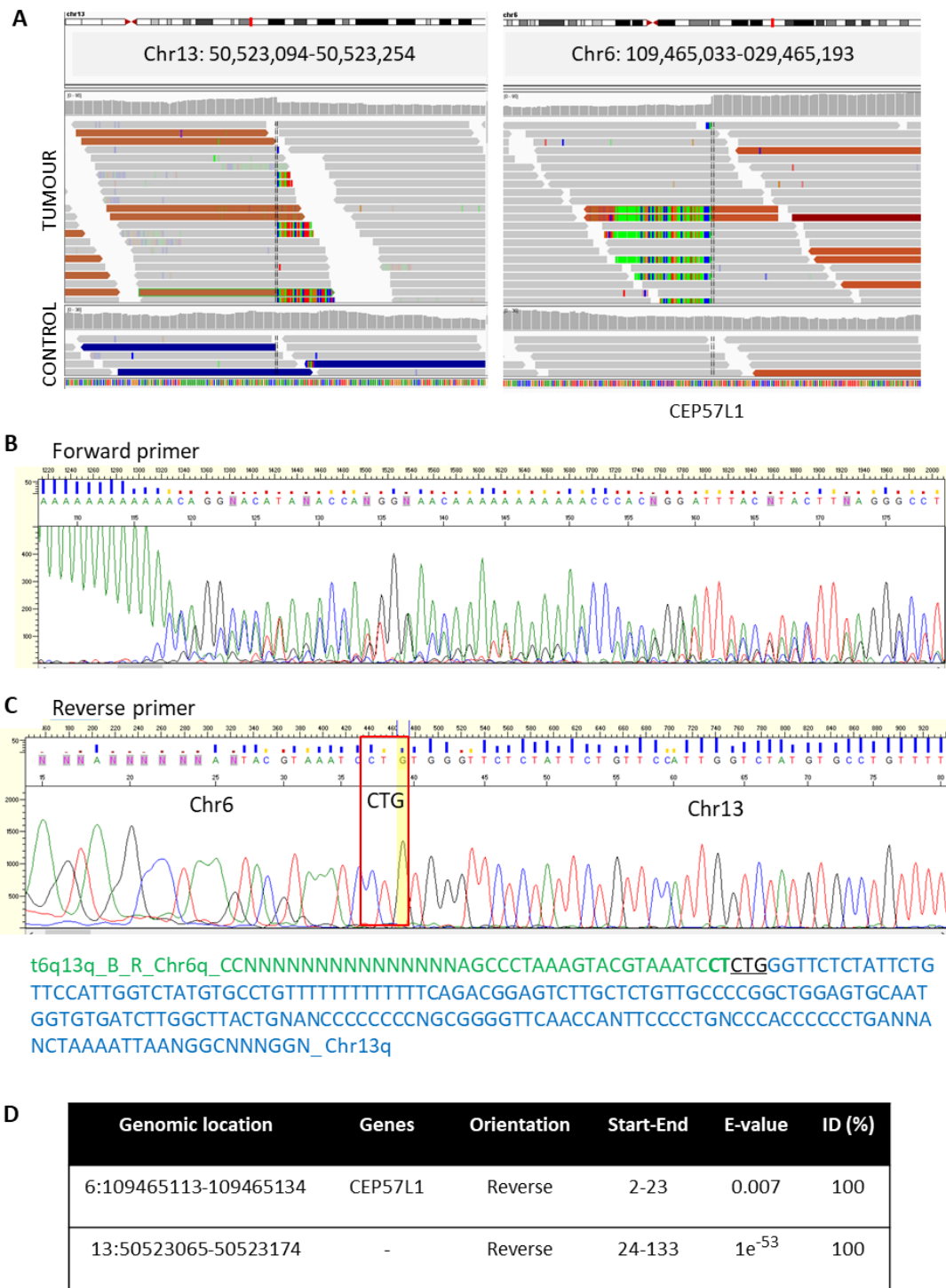
Supplementary Figure 5. Copy number deletion at chr6q21 and chr13q14.

Ensembl snapshot image of the deleted loci at (A) chr6q21 and at (B) chr13q14.2-13q14.3. **Gene Legend:** Ensembl protein coding, merged Ensembl/Havana, processed transcript, pseudogene, RNA gene. Detected in DB17 CLL-B cells.



Supplementary Figure 6. Sequencing the breakpoint A of the 6q:13q translocation.

(A) Visualisation in *IGHV* of the 6q:13q translocation fusion point A (Chr6:107,974,284-Chr13:51,501,490). (B) Sanger sequencing trace obtained with the forward and (C) reverse primers after sequencing breakpoint A. (D) Results obtained after BLAST-alignment of the sequence in Ensembl GRCh37. Microhomology in black and highlighted. Detected in DB17 CLL-B cells.



Supplementary Figure 7. Sequencing the breakpoint B of the 6q:13q translocation.

(A) Visualisation in *IGHV* of the 6q:13q translocation fusion point B (Chr13:50,523,174-Chr6:109,465,113). (B) Sanger sequencing trace obtained with the forward and (C) reverse primers after sequencing breakpoint B. (D) Results obtained after BLAST-alignment of the sequence in Ensembl GRCh37. Microhomology in black and highlighted. Detected in DB17 CLL-B cells.

SUPPLEMENTARY TABLES

Supplementary Table 1. List of oligonucleotides used during this study.

APPLICATION	PRIMER NAME	SEQUENCE (5'-3')
Development of STELA and the telomere fusion assay at the 5p chromosome end	5p4_F	AGGGAGTGCATTAGCATACAGGTG
	5p5_R	GGAGCAGCATTCTCTCACCACAG
	5p6_1.4_F	CGTAGAGGAGGGTGAACCTC
	5p7_2_R	AGCACGGATAAGGAGGACATTAAC
	5p8_3.2_R	CCTCTACTAACCTTTAAGGCTGTG
	5p9_6_R	AGATATTTCCCATGTAGCCGCAAC
Characterisation of the 1.4Kb 5p8 background band (non-fusion amplicon)	5p_seq1	AGCTGATTGCAGGAGCTTTC
	5p_seq2	CCTTATAACGGGACATAATTGGA
	5p8.2_R	AAGGCTGTGAACCCTGTAATCTAG
	5p8.3_R	ATCTAGGTATCAGGCTGGCTTTTC
Measurement of the telomere length at different chromosome ends using STELA	5p5	GGAGCAGCATTCTCTCACCACAG
	XpYpE2	TTGTCTCAGGGTCCTAGTG
	17pseq1rev	GAATCCACGGATTGCTTTGTGTAC
	18qrev4M	CACAGGGATGGTTAGGTATCTC
	2p2	GAGCTGCGTTTTGCTGAGCAC
	12q -197A	GGGAGATCCACACCGTAGCA
	Teltail	TGCTCCGTGCATCTGGCATC
	Tel2	TGCTCCGTGCATCTGGCATCTAACCT
Measurement of the Telomere Variant Repeat content using TVR assay	5p5	GGAGCAGCATTCTCTCACCACAG
	XpYpE2	TTGTCTCAGGGTCCTAGTG
	17pseq1rev	GAATCCACGGATTGCTTTGTGTAC
	TAG-TelW	TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCT NACCCTA
	TAG-TelX	TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCT NACCCTC
	TAG-TelY	TCATGCGTCCATGGTCCGGACCCTTACCCTTACCCT NACCCTG
Detection of telomere fusions using the 5p8:17p6:XpYpM:16p1:21q1 telomere fusion assay	5p8	CCTCTACTAACCTTTAAGGCTGTG
	17p6	GGCTGAACTATAGCCTCTGC
	XpYpM	ACCAGGTTTTCCAGTGTGTT
	16p1	TGGACTTCTCACTTCTAGGGCAG

	21q1	CTTGGTGTCTCGAGAGAGGTAG
Validation of mutations by Sanger sequencing SNVs identified from WGS of a CLL patient sample	BRAF_F	AAGATGGAGTTTTGCCCTTGTAC
	BRAF_R	CAAGATCCCCAAATGATTCTGTGTC
	ATR_F	AAGAATTAGCTGGGTATGGTGGTG
	ATR_R	ATGTGGTATGCTCTCACTCCATG
	EHMT1_F	GGGAAGGAGTAGATGTTTTCCAGT
	EHMT1_R	GTTCTTACCGCCACTAATATGCTG
	IGF1R_F	GTGGGGAAAGGAGAGGTGTATATG
	IGF1R_R	AGCAGTACACCAGTCAAGTTTCTC
	TMPRSS15_F	CTAAAGCTGAGTTCCTGGAGTAGC
	TMPRSS15_R	TTGGACCTGTTGCCTGTTTCATTC
	POT1_F	ATGCCCCAGTTCTCTTAGTGATG
	POT1_R	ACTGGTCTCTTCTCTGTCTATC
	TP53_F	AGAGCAATCAGTGAGGAATCAGAG
	TP53_R	CTTTATCTGTTCACTTGTGCCCTG
	REVEL_F	CACCAAATGCTGAAGGTGTTGAC
	REVEL_R	TCAAAGCTGGACCAAGCATATACC
CREBBP_F	ACTTATAGACTCGGGATGTCTTGC	
CREBBP_R	GAGTAAAATCAGTTGCGTCAGAGG	
Sanger sequencing the fusion points A and B of a 6q:13q translocation identified from WGS of a CLL patient sample	T6Q13Q_A_F	GCTCACTTTATTGCAGTGGTCTAG
	T6Q13Q_A_R	CAGAAAGAGCAAGGACACAAAGAC
	T6Q13Q_B_F	CACGTGCCTATAATCTCAGCTACT
	T6Q13Q_B_R	GAGATCAACATTAGTCACAGCTGG
Validation of POT1 mutations in a cohort of CLL patient samples by Sanger sequencing	POT1A	CCACAGAGTACATATATGTTAG
	POT1B	TAAGGAATAGCAGACTCATATG
	POT2A	ACCTACGTAACAAACCTCCATG
	POT2B	CTCACACGTTATTTAATAGGACTG
	POT3A	ACTTTATATGTACTAGGTTGTCTG
	POT3B	CATGGAGTGACACTTAATAGAG

Supplementary Table 2. Characteristics of 5p STELA PCR primers.

DISTANCE TO TELOMERE (Kb)	ID	PRIMER SEQUENCE 5'-3'	LENGTH (bp)	T _m (°C)	GC (%)
0.056	5p4_R	AGGGAGTGCATTAGCATACAGGTG	24	62.50	50.00
0.367	5p5_F	GGAGCAGCATTCTTCACCACAG	24	63.70	54.17
Product size= 0.335Kb					

Supplementary Table 3. Comparison of the mean telomere length (Kb) and standard deviation at 5p, 17p and XpYp for 57 CLL patient samples.

SOURCE	Thesis #	5p Mean	5p SD	17p Mean	17p SD	XpYp Mean	XpYp SD
LRF CLL4	1	1.57	1.28	1.43	1.44	1.22	1.32
LRF CLL4	2	3.34	1.59	3.24	1.77	2.78	0.76
LRF CLL4	3	3.71	2.09	2.27	1.06	2.37	1.05
LRF CLL4	4	3.15	0.85	3.46	1.25	2.82	1.01
LRF CLL4	5	1.96	1.82	1.84	1.22	0.88	0.84
LRF CLL4	6	2.66	0.66	2.46	0.64	2.71	0.95
LRF CLL4	8	3.69	2.31	3.86	1.47	2.71	0.99
LRF CLL4	9	1.38	0.64	1.64	1.05	1.40	0.42
LRF CLL4	10	1.59	1.42	1.42	0.47	1.21	1.09
LRF CLL4	11	1.95	0.74	2.14	0.94	1.88	0.76
LRF CLL4	12	2.63	0.71	3.06	1.20	2.55	0.84
LRF CLL4	13	3.45	0.72	4.16	1.63	2.85	0.86
LRF CLL4	14	1.59	0.76	2.52	2.16	1.97	1.53
LRF CLL4	15	3.13	1.84	2.84	1.18	2.68	1.09
LRF CLL4	16	2.03	1.62	2.87	2.22	2.80	3.48
LRF CLL4	17	2.59	0.93	2.81	1.11	2.21	1.29
LRF CLL4	18	1.66	1.09	1.60	1.17	1.63	0.93
LRF CLL4	19	4.02	1.22	3.36	1.51	3.28	1.47
LRF CLL4	20	2.77	1.3	3.49	0.83	2.11	0.67
LRF CLL4	21	2.58	0.76	2.62	0.49	2.23	0.87
LRF CLL4	22	3.33	1.66	2.71	0.77	2.25	0.67
LRF CLL4	23	1.84	2.03	1.67	1.03	1.19	0.48
LRF CLL4	24	3.44	0.67	2.70	0.76	2.43	0.76
LRF CLL4	25	2.43	1.85	1.71	1.56	2.16	1.76
LRF CLL4	26	4.40	2.08	3.80	0.96	3.04	0.82
LRF CLL4	27	4.49	3.21	2.68	0.61	2.04	0.52
LRF CLL4	28	3.70	1.95	2.84	1.06	2.31	0.78
LRF CLL4	29	3.29	1.38	2.84	0.81	2.28	0.51
LRF CLL4	30	2.95	0.85	2.94	0.93	1.93	0.78
LRF CLL4	31	2.94	1.78	2.15	1.00	3.01	0.80
LRF CLL4	32	3.09	1.4	2.56	0.82	2.05	0.59

LRF CLL4	33	4.13	1.12	3.44	1.07	2.69	0.99
LRF CLL4	34	3.47	1.1	3.34	0.85	2.25	1.23
LRF CLL4	35	1.49	1.75	1.07	0.49	1.61	0.76
LRF CLL4	36	3.48	0.9	2.88	1.16	2.40	1.13
LRF CLL4	37	3.58	1.56	4.41	1.58	3.43	0.83
HOSPITAL	41	4.80	0.83	2.30	0.98	2.63	1.47
HOSPITAL	42	3.05	0.84	2.70	1.03	2.32	0.82
HOSPITAL	43	2.52	1.10	2.70	1.06	1.76	0.63
HOSPITAL	45	2.90	1.87	3.42	1.34	3.10	0.90
HOSPITAL	46	5.04	2.00	3.21	1.19	4.13	1.55
HOSPITAL	47	3.66	2.06	4.72	1.52	3.28	1.21
HOSPITAL	48	3.78	1.84	4.03	2.12	3.87	1.32
HOSPITAL	51	1.97	1.84	2.64	0.75	1.39	0.84
HOSPITAL	52	3.95	1.94	2.97	0.99	2.72	0.70
HOSPITAL	53	6.53	2.33	5.72	2.52	5.36	1.41
HOSPITAL	54	5.42	2.23	5.26	1.54	5.36	0.93
HOSPITAL	55	3.41	1.60	2.92	1.30	2.33	0.91
HOSPITAL	56	4.92	1.56	3.69	1.17	3.15	1.06
HOSPITAL	65	6.62	3.16	6.75	2.83	5.59	2.36
HOSPITAL	70	2.25	1.34	1.76	1.01	2.29	1.40
ADMIRE	105	1.93	1.90	1.85	1.76	1.04	0.54
ADMIRE	108	3.32	2.24	3.77	2.79	2.34	1.63
ADMIRE	112	2.25	2.17	1.81	1.40	1.74	1.32
ADMIRE	118	2.12	1.73	1.67	1.35	1.34	1.13
ADMIRE	226	3.40	1.53	2.66	0.98	3.05	1.18
ADMIRE	251	3.92	3.02	2.90	1.39	2.96	1.68
TOTAL		3.18	1.16	2.92	1.07	2.51	0.98

Supplementary Table 4. TL before and after correcting for TVR limit (Kb).

ID #	STELA TL		TVR limit		TL corrected by TVR	
	5p	XpYp	5p	XpYp	5p	XpYp
48	3.41	3.46	2.24	0.73	1.17	2.73
53	6.16	4.95	1.47	0.49	4.69	4.46
47	3.29	2.87	1.99	0.86	1.30	2.01
52	3.58	2.31	1.39	0.85	2.19	1.46
54	5.05	4.95	2.14	3.05	2.91	1.90
40	2.97	1.01	0.94	0.50	2.03	0.51
41	4.43	2.23	2.11	0.45	2.32	1.78
42	2.68	1.91	1.40	1.53	1.28	0.38
56	4.55	2.74	3.09	1.70	1.46	1.04
46	4.67	3.73	1.05	0.34	3.62	3.39
TOTAL	4.08	3.02	1.78	1.05	2.30	1.97

Supplementary Table 5. Oligonucleotide sequences for the 5p chromosome end: distance to telomere, primer ID, sequence, length, expected T_m and GC content.

DISTANCE TO TELOMERE (Kb)	ID	PRIMER SEQUENCE 5'-3'	LENGTH (bp)	T _m (°C)	GC (%)
0.056	5p4_R	AGGGAGTGCATTAGCATACAGGTG	24	62.50	50.00
0.367	5p5_F	GGAGCAGCATTCTCTTCACCACAG	24	63.70	54.17
1.310	5p6_R	CGTAGAGGAGGGTGGAACCTC	21	61.30	61.90
1.925	5p7_F	AGCACGGATAAGGAGGACATTAAC	24	60.44	45.83
3.406	5p8_F	CCTCTACTAACCTTTAAGGCTGTG	24	58.58	45.83
5.930	5p9_F	AGATATTTCCCATGTAGCCGCAAC	24	61.28	45.83

Supplementary Table 6. 5p primer combination and PCR product.

	F primer	R primer	PCR product (Kb)	PCR amplification	Optimal T _m
A	5p5	5p4	0.335	Yes	62-65°C
B	5p7	5p4	1.893	No	-
C	5p8	5p4	3.374	No	-
D	5p9	5p4	5.898	No	-
E	5p7	5p6	0.636	Yes	62-65°C
F	5p8	5p6	2.117	Yes	62°C
G	5p9	5p6	4.641	Yes	62-65°C

Supplementary Table 7. Telomere dynamics at the 5p chromosome end for HCT116^{DN-hTERT}.

	HCT116 ^{DN-hTERT} WT c11					HCT116 ^{DN-hTERT} LIG3 ^{-/-} -mL3+nuc. LIG3 c15				
Day	10	42	49	66	91	22	103	153	173	224
Mean 5p TL (Kb)	2.76	1.78	1.91	1.49	1.64	1.25	1.70	1.88	2.36	3.17
Erosion (Kb)	1.27									
Elongation (Kb)				0.15		1.92				

Supplementary Table 8. Number and proportion of CLL patients for each category of fusions.

CATEGORY	#FUSIONS	FREQUENCY	ALL				HOSPITAL				LRF CLL4				ARCTIC and ADMIRE			
			#	%	#	%	#	%	#	%	#	%	#	%	#	%		
NONE	0	0.00E+00	78	28.26	78	28.26	11	33.33	11	33.33	14	35.90	14	35.90	53	25.98	53	25.98
LOW	1	6.00E-06	72	26.09	143	51.81	4	12.12	11	33.33	13	33.33	20	51.28	55	26.96	112	54.90
	2	1.20E-05	41	14.86			3	9.09			5	12.82			33	16.18		
	3	1.80E-05	30	10.87			4	12.12			2	5.13			24	11.76		
MEDIUM	4	2.40E-05	22	7.97	40	14.49	5	15.15	8	24.24	2	5.13	4	10.26	15	7.35	28	13.73
	5	3.00E-05	13	4.71			1	3.03			2	5.13			10	4.90		
	6	3.60E-05	5	1.81			2	6.06			0	0.00			3	1.47		
HIGH	>7	4.20E-05	15	5.43	15	5.43	3	9.09	3	9.09	1	2.56	1	2.56	11	5.39	11	5.39
TOTAL			276	100	276	100	33	100	33	100	39	100	39	100	204	100	204	100

*Number of telomere fusion events and frequency of fusions detected from 10 fusion PCR reactions, 100ng gDNA/reaction.

Supplementary Table 9. Description of potential 5p fusions detected in CLL patient samples.

CATEGORY	#FUSIONS	FREQUENCY	5P FUSIONS: TYPE & SIZE (Kb)	ID-CLL
LOW	1	6.00E ⁻⁰⁶	5p-21q (~2.75)	68
LOW	1	6.00E ⁻⁰⁶	5p-5p (~0.8)	29
LOW	1	6.00E ⁻⁰⁶	5p-16p/21q (~0.5)	144
LOW	1	6.00E ⁻⁰⁶	5p-16p/21q (~2.6)	182
LOW	1	6.00E ⁻⁰⁶	5p-5p (~1.6)	198
LOW	2	1.20E ⁻⁰⁵	5p-16p/21q (~2.75)	107
LOW	2	1.20E ⁻⁰⁵	5p-5p (~1.75)	86
LOW	2	1.20E ⁻⁰⁵	5p-5p (2kb)	142
LOW	2	1.20E ⁻⁰⁵	5p-5p (~1.75)	153
LOW	2	1.20E ⁻⁰⁵	5p-5p (1.75)	160
LOW	2	1.20E ⁻⁰⁵	5p-5p (~1.75); 5p-5p (~4.1)	269
LOW	2	1.20E ⁻⁰⁵	5p-5p (~1.75); 5p-16p/21q (~2.75)	274
LOW	3	1.80E ⁻⁰⁵	5p-16p/21q (~2.75)	236
LOW	3	1.80E ⁻⁰⁵	5p-5p (~1.75)	76
LOW	3	1.80E ⁻⁰⁵	5p-5p (~2.75); 5p-5p (~2.9)	117
MEDIUM	4	2.40E ⁻⁰⁵	5p-16p/21q (~2.75) x2	60
MEDIUM	4	2.40E ⁻⁰⁵	5p-16p/21q (~2.75)	217
MEDIUM	4	2.40E ⁻⁰⁵	5p-5p (~1.6)	232
MEDIUM	4	2.40E ⁻⁰⁵	5p-5p (~1.75)	233
MEDIUM	4	2.40E ⁻⁰⁵	5p-5p (~0.25); 5p-5p (~1.2)	243
MEDIUM	4	2.40E ⁻⁰⁵	5p-5p (~1.75)	255
MEDIUM	4	2.40E ⁻⁰⁵	5p-5p (~0.75)	271
MEDIUM	5	3.00E ⁻⁰⁵	5p (~2.75)	113
MEDIUM	5	3.00E ⁻⁰⁵	5p-16p/21q (~2.75)	120
MEDIUM	5	3.00E ⁻⁰⁵	5p-5p (~3.7); 5p-5p (~1.2)	235
MEDIUM	5	3.00E ⁻⁰⁵	5p-5p (~0.75)	246
MEDIUM	5	3.00E ⁻⁰⁵	5p-17/XpYp (~2.0)	272
MEDIUM	6	3.60E ⁻⁰⁵	5p-16p (~0.4)	59
MEDIUM	6	3.60E ⁻⁰⁵	5p-17/XpYp (~3.5)	112
HIGH	7	4.20E ⁻⁰⁵	5p-16p/21q (~2.75)	219
HIGH	7	4.20E ⁻⁰⁵	5p-17/XpYp (~6.5)	253
HIGH	8	4.80E ⁻⁰⁵	5p-5p (~0.5)	108
HIGH	8	4.80E ⁻⁰⁵	5p-5p (~0.25); 5p-16p/21q (~2.75)	121
HIGH	8	4.80E ⁻⁰⁵	5p-5p (~1.75); 5p-16p/21q (~1.0)	215
HIGH	8	4.80E ⁻⁰⁵	5p-5p (~0.75)	251
HIGH	10	6.00E ⁻⁰⁵	5p-21q (~2.75); 5p-5p (~4kb)	65
HIGH	12	7.20E ⁻⁰⁵	5p-5p (~1.75)	226
HIGH	~20	~1.20E ⁻⁰⁴	5p-YpYp (~2.5)	41
HIGH	~20	~1.20E ⁻⁰⁴	5p-21q (~5.5); 5p-21q (~7.5)	70
HIGH	~20	~1.20E ⁻⁰⁴	5p-5p (~1.6); 5p-5p (~2.3); 5p-16p/21q (~2.75)	118

Supplementary Table 10. Summary of events resulting from intra-chromosomal fusion analysis.

INTRA	TOTAL	"	U	T/R	Tel-Tel (#00)	Subtel- Tel (#0)	Intra (#1)	Intra/Inter (#1o2)	Inter (#2T)	2q13 (#2V)	Genomic (#2G)	Complex (#2C)	Mt DNA (#2M)
DB17	47	0	6	41	0	36	5	0	0	0	0	0	0
DB59	49	1	3	45	3	38	2	0	1	0	0	1	0
DB60	10	0	0	10	0	3	5	0	1	0	0	1	0
DB61	7	0	1	6	0	6	0	0	0	0	0	0	0
DB62	19	0	1	18	4	8	3	0	3	0	0	0	0
DB63	58	0	9	49	3	41	5	0	0	0	0	0	0
DB64	8	0	1	7	0	1	6	0	0	0	0	0	0
DB65	5	0	1	4	3	0	1	0	0	0	0	0	0
DB66	7	0	1	6	0	2	3	0	1	0	0	0	0
TOTAL	210	1	23	186	13	135	30	0	6	0	0	2	0
5p	68	0	13	55	4	41	7	0	3	0	0	0	0
17p	65	1	1	63	5	43	13	0	1	0	0	1	0
XpYp	77	0	9	68	4	51	10	0	2	0	0	1	0

Supplementary Table 11. Summary of events resulting from inter-chromosomal fusion analysis.

INTER	TOTAL	"	U	T/R	Tel-Tel (#0D)	Subtel- Tel (#0)	Intra (#1)	Intra/Inter (#1o2)	Inter (#2T)	2q13 (#2V)	Genomic (#2G)	Complex (#2C)	Mt DNA (#2M)
DB17	131	3	9	119	0	52	10	21	29	2	5	0	0
DB59	116	7	18	91	0	58	1	3	18	1	9	1	0
DB60	148	7	9	132	0	48	0	9	63	1	8	3	0
DB61	53	6	6	41	0	11	0	1	15	1	12	0	1
DB62	173	12	13	148	0	28	2	4	108	1	4	1	0
DB63	29	5	3	21	1	13	1	2	3	0	1	0	0
DB64	41	8	7	26	1	5	1	4	6	0	9	0	0
DB65	41	6	7	28	1	2	0	0	4	3	17	0	1
DB66	149	14	13	122	0	8	4	5	91	2	10	0	2
TOTAL	881	68	85	728	3	225	19	49	337	11	75	5	4
TOTAL (≥3 reads)	806	68	69	669	3	221	18	49	326	11	36	5	0
TOTAL (<3 reads)	75	0	16	59	0	4	1	0	11	0	39	0	4

Supplementary Table 12. Summary of events resulting from both intra- and inter-chromosomal fusion analyses.

ALL	TOTAL	"	U	T/R	Tel-Tel (00)	Subtel- Tel (0)	Intra (1)	Intra/Inter (1o2)	Inter (2T)	2q13 (2V)	Genomic (2G)	Complex (2C)	Mt DNA (2M)
INTER (≥ 3 reads)	806	68	69	669	3	221	18	49	326	11	36	5	0
INTER (< 3 reads)	75	0	16	59	0	4	1	0	11	0	39	0	4
INTER-TOTAL	881	68	85	728	3	225	19	49	337	11	75	5	4
INTRA (≥ 3 reads)	210	1	23	186	13	135	30	0	6	0	0	2	0
INTRA (< 3 reads)	0	0	0	0	0	0	0	0	0	0	0	0	0
INTRA-TOTAL	210	0	23	187	13	135	31	0	6	0	0	2	0
TOTAL (≥ 3 reads)	1016	68	92	856	16	356	49	49	332	11	36	7	0
TOTAL (< 3 reads)	75	0	16	59	0	4	1	0	11	0	39	0	4
TOTAL	1091	68	108	914	16	360	49	49	343	11	75	7	4

Table indicating the number of events detected with ≥ 3 or < 3 reads.

Supplementary Table 13. Type and proportion of TVRs identified.

Abundance (Lee, 2014)	TVR	TOTAL	TOTAL%
1	TTAGGG	1483	76.05
6	TTGGGG	88	4.51
5	GTAGGG	84	4.31
3	TGAGGG	71	3.64
2	TCAGGG	59	3.03
	TTAGGGG	57	2.92
4	TTCGGG	36	1.85
	TTTAGGG	20	1.03
9	CTAGGG	19	0.97
7	TAAGGG	6	0.31
	TTAAGGG	6	0.31
	TTAGGGGGG	5	0.26
8	ATAGGG	4	0.21
	TTAGTG	4	0.21
	GTGGGG	3	0.15
10	TTTGGG	2	0.10
	TTAGCG	1	0.05
	TGGGGG	1	0.05
1	CTGGGG	1	0.05
	TOTAL	1950	100.00

Supplementary Table 14. Biological process gene ontologies enriched in gene list, with validated fusion junction, disrupted by inter-chromosomal fusion events.

DAVID assigned GO Term Biological Process	# genes within GO category	% genes within GO category	<i>p</i>-value	Benjamini-Hochberg corrected <i>p</i>-value
biological regulation	22	68.8	7.10E-02	9.90E-01
<i>HTR7, CD8A, CSMD1, POLDIP3, FGGY, FTO, NOX5, RORA, SHQ1, TBC1D15, ADGRL1, DGKB, DMD, EVI5, ECE1, KIF26B, NTF3, PTC3, SLC30A10, TESPA1, VPS13D, ZNF254</i>				
anatomical structure development	13	40.6	7.90E-02	9.80E-01
<i>CD8A, FTO, ND4, NOX5, RORA, TBC1D15, ADGRL1, DMD, EVI5, ECE1, KIF26B, NTF3, TESPA1</i>				
developmental process	13	40.6	9.50E-02	9.60E-01
<i>CD8A, FTO, ND4, NOX5, RORA, TBC1D15, ADGRL1, DMD, EVI5, ECE1, KIF26B, NTF3, TESPA1</i>				
regulation of biological quality	11	34.4	2.50E-02	9.80E-01
<i>HTR7, CSMD1, FGGY, FTO, RORA, ADGRL1, DGKB, DMD, ECE1, SLC3A10</i>				
system development	11	34.4	8.10E-02	9.80E-01
<i>CD8A, FTO, ND4, NOX5, RORA, ADGRL1, DMD, ECE1, KIF26B, NTF3, TESPA1</i>				
cellular component assembly	9	28.1	3.10E-02	9.90E-01
<i>ND4, ND5, SHQ1, TBC1D15, ADGRL1, DMD, EVI5, KIF13A, TESPA1</i>				
macromolecule localization	9	28.1	4.60E-02	9.90E-01
<i>POLDIP3, NOX5, SHQ1, TBC1D15, DMD, EVI5, KIF13A, KIF26B, VPS13D</i>				
establishment of localization in cell	8	25	2.40E-02	9.90E-01
<i>POLDIP3, TBC1D15, ADGRL1, DMD, EVI5, KIF13A, KIF26B, VPS13D</i>				
regulation of localization	8	25	6.40E-02	1.00E+00
<i>FTO, NOX5, SHQ1, TBC1D15, DMD, EVI5, NTF3, SLC30A10</i>				
cellular localization	8	25	7.20E-02	9.90E-01
<i>POLDIP3, TBC1D15, ADGRL1, DMD, EVI5, KIF13A, KIF26B, VPS13D</i>				
anatomical structure morphogenesis	8	25	8.40E-02	9.70E-01
<i>NOX5, RORA, TBC1D15, DMD, EVI5, ECE1, KIF26B, NTF3</i>				
regulation of multicellular organismal process	8	25	8.80E-02	9.70E-01
<i>FTO, NOX5, RORA, ADGRL1, DMD, ECE1, NTF3, TESPA1</i>				
regulation of developmental process	7	21.9	8.80E-02	9.70E-01
<i>FTO, TBC1D15, ADGRL1, DMD, EVI5, NTF3, TESPA1</i>				
cellular macromolecular complex assembly	6	18.8	1.50E-02	1.00E+00
<i>ND4, ND5, SHQ1, DMD, KIF13A, TESPA1</i>				
intracellular transport	6	18.8	8.20E-02	9.70E-01
<i>POLDIP3, TBC1D15, EVI5, KIF13A, KIF26B, VPS13D</i>				
homeostatic process	6	18.8	9.10E-02	9.60E-01
<i>CSMD1, FGGY, FTO, RORA, DMD, SLC30A10</i>				
protein complex subunit organization	6	18.8	9.70E-02	9.60E-01
<i>ND4, ND5, DMD, KIF13A, PTC3, TESPA1</i>				

cellular protein complex assembly	5	15.6	9.70E-03	1.00E+00
<i>ND4, ND5, DMD, KIF13A</i>				
cell-cell adhesion	5	15.6	9.10E-02	9.60E-01
<i>CD8A, RORA, ADGRL1, KIF26B, TESPA1</i>				
single organismal cell-cell adhesion	4	12.5	9.10E-02	9.60E-01
<i>CD8A, RORA, KIF26B, TESPA1</i>				
activation of GTPase activity	3	9.4	6.10E-03	1.00E+00
<i>TBC1D15, EVI5, NTF3</i>				
calcium-mediated signalling	3	9.4	2.00E-02	1.00E+00
<i>CD8A, ADGRL1, DMD</i>				
receptor metabolic process	3	9.4	2.10E-02	9.90E-01
<i>DMD, ECE1, NTF3</i>				
lymphocyte differentiation	3	9.4	3.70E-02	9.90E-01
<i>CD8A, RORA, TESPA1</i>				
second-messenger-mediated signalling	3	9.4	4.70E-02	9.90E-01
<i>CD8A, ADGRL1, DMD</i>				
response to hypoxia	3	9.4	6.70E-02	1.00E+00
<i>ND4, ND5, RORA</i>				
vacuolar transport	3	9.4	6.80E-02	9.90E-01
<i>EVI5, KIF13A, VPS13D</i>				
response to decreased oxygen levels	3	9.4	7.10E-02	9.90E-01
<i>ND4, ND5, RORA</i>				
mitochondrial electron transport, NADH to ubiquinone	2	6.2	7.60E-02	9.90E-01
<i>ND4, ND5</i>				
regulation of cilium assembly	2	6.2	8.70E-02	9.70E-01
<i>TBC1D15, EVI5</i>				
mitochondrial respiratory chain complex I assembly and biogenesis	2	6.2	9.00E-02	9.70E-01
<i>ND4, ND5</i>				
regulation of vesicle fusion	2	6.2	9.00E-02	9.70E-01
<i>TBC1D15</i>				

Supplementary Table 15. List of human genes with validated fusion junction submitted to DAVID and GSEA.

ENSEMBL	ENTREZ	Gene	Gene Name
ENSG00000048707	55187	<i>VPS13D</i>	vacuolar protein sorting 13 homolog D
ENSG00000067208	7813	<i>EVI5</i>	ecotropic viral integration site 5
ENSG00000069667	6095	<i>RORA</i>	RAR related orphan receptor A
ENSG00000072071	22859	<i>LPHN1</i> (<i>ADGRL1</i>)	Latrophilin (adhesion G protein-coupled receptor L1)
ENSG00000088205	8886	<i>DDX18</i>	DEAD-box helicase 18
ENSG00000100227	84271	<i>POLDIP3</i>	DNA polymerase delta interacting protein 3
ENSG00000111906	51020	<i>HDCC2</i>	HD domain containing 2
ENSG00000117298	1889	<i>ECE1</i>	endothelin converting enzyme 1
ENSG00000121749	64786	<i>TBC1D15</i>	TBC1 domain family member 15
ENSG00000132300	55037	<i>PTCD3</i>	pentatricopeptide repeat domain 3
ENSG00000135426	9840	<i>TESPA1</i>	thymocyte expressed, positive selection associated 1
ENSG00000135749	80003	<i>PCNX2</i>	pecanex homolog 2 (Drosophila)
ENSG00000136267	1607	<i>DGKB</i>	diacylglycerol kinase beta
ENSG00000137177	63971	<i>KIF13A</i>	kinesin family member 13A
ENSG00000140718	79068	<i>FTO</i>	FTO, alpha-ketoglutarate dependent dioxygenase
ENSG00000144736	55164	<i>SHQ1</i>	SHQ1, H/ACA ribonucleoprotein assembly factor
ENSG00000146521	26238	<i>C6orf123</i> (<i>LINC01558</i>)	chromosome 6 open reading frame 123/ long intergenic non-protein coding RNA 1558
ENSG00000148680	3363	<i>HTR7</i>	5-hydroxytryptamine receptor 7
ENSG00000153563	925	<i>CD8A</i>	CD8a molecule
ENSG00000162849	55083	<i>KIF26B</i>	kinesin family member 26B
ENSG00000165626	222389	<i>BEND7</i>	BEN domain containing 7
ENSG00000172456	55277	<i>FGGY</i>	FGGY carbohydrate kinase domain containing
ENSG00000183117	64478	<i>CSMD1</i>	CUB and Sushi multiple domains 1
ENSG00000185652	4908	<i>NTF3</i>	neurotrophin 3
ENSG00000188859	149297	<i>FAM78B</i>	family with sequence similarity 78 member B
ENSG00000196660	55532	<i>SLC30A10</i>	solute carrier family 30 member 10
ENSG00000198786	4540	<i>MT-ND5</i>	NADH dehydrogenase, subunit 5 (complex I)
ENSG00000198886	4538	<i>MT-ND4</i>	NADH dehydrogenase, subunit 4 (complex I)
ENSG00000198947	1756	<i>DMD</i>	dystrophin
ENSG00000213096	9534	<i>ZNF254</i>	zinc finger protein 254
ENSG00000255346	79400	<i>NOX5</i>	NADPH oxidase 5
ENSG00000231473		<i>LINC00441</i>	long intergenic non-protein coding RNA 441
ENSG00000251680		<i>CTC-575N7.1</i>	novel antisense; non protein coding
ENSG00000262987		<i>RP11-520H11.1</i>	unprocessed pseudogene

Supplementary Table 16. GSEA MSigDB gene set collections (Mootha, 2003; Subramanian, 2005).

MSigDB gene set collections	
H	Hallmark gene sets are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.
C1	Positional gene sets for each human chromosome and cytogenetic band.
C2	Curated gene sets from online pathway databases, publications in PubMed, and knowledge of domain experts.
C3	Motif gene sets based on conserved cis-regulatory motifs from a comparative analysis of the human, mouse, rat, and dog genomes.
C4	Computational gene sets defined by mining large collections of cancer-oriented microarray data.
C5	GO gene sets consist of genes annotated by the same GO terms.
C6	Oncogenic signatures defined directly from microarray gene expression data from cancer gene perturbations.
C7	Immunologic signatures defined directly from microarray gene expression data from immunologic studies.
http://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/MSigDB_collections	

Supplementary Table 17. Enrichment in GSEA MSig collections from a list of disrupted genes with validated fusion junction.

Collections	# Overlaps Shown	# Gene Sets in Collections	# Genes in Comparison (n)	# Genes in Universe (N)
C2	1	4729	31	45956
C3	2	836	31	45956
C4	5	858	31	45956
C6	2	189	31	45956

Supplementary Table 18. Gene set overlaps found between a list of genes with validated fusion junction disrupted by inter-chromosomal telomere fusions and the Molecular Signatures Database (MSigDB v. 5.2) using Gene Set Enrichment Analysis (GSEA v.5.0; Broad Institute).

MSigDB collection	MSigDB gene set name	Number of genes in gene set (K)	Description of gene set	Number of gene list genes in overlap (k)	k/K	p-value	False Discovery Rate (FDR) FDR q-value
C2	<u>PEPPER_CHRONIC_LYMPHOCYTIC_LEUKAEMIA_UP</u>	33	Genes up-regulated in CD38+ [GeneID=952] CLL (chronic lymphocytic leukaemia) cells.	3 (HTR7, LPHN1, KIF26B)	0.091	1.5 e ⁻⁶	7.07 e ⁻³
C3	<u>V\$HNF1_Q6</u>	253	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif WRGTTAATNATTAACNNN which matches annotation for TCF1: transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor	4 (DMD, RORA, NTF3, HTR7)	0.016	2.51 e ⁻⁵	1.31 e ⁻²
C3	<u>V\$OCT_C</u>	268	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CTNATTTGCATAY. Motif does not match any known transcription factor	4 (DMD, LPHN1, KIF13A, POLDIP3)	0.015	3.14 e ⁻⁵	1.31 e ⁻²
C4	<u>MODULE_67</u>	230	Genes in the cancer module 67 (Breast and liver cancer).	4 (HDCC2, NTF3, KIF26B, VPS13D)	0.017	1.73 e ⁻⁵	1.48 e ⁻²
C4	<u>MODULE_279</u>	141	Genes in the cancer module 279 (Breast cancer).	3 (DDX18, DMD, KIF13A)	0.021	1.19 e ⁻⁴	4.63 e ⁻²
C4	<u>MODULE_334</u>	166	Genes in the cancer module 334 (Breast cancer).	3 (DDX18, DMD, KIF13A)	0.018	1.93 e ⁻⁴	4.63 e ⁻²
C4	<u>GCM_CSNK1D</u>	32	Neighbourhood of CSNK1D	2 (EVI5, POLDIP3)	0.062	2.16 e ⁻⁴	4.63 e ⁻²
C4	<u>GCM_CSNK1A1</u>	36	Neighbourhood of CSNK1A1	2 (TBC1D15, POLDIP3)	0.055	2.73 e ⁻⁴	4.69 e ⁻²
C6	<u>KRAS.BREAST_UP.V1_DOWN</u>	145	Genes down-regulated in epithelial breast cancer cell lines over-expressing an oncogenic form of KRAS [Gene ID=3845] gene.	3 (NTF3, FOX5, FGGY)	0.021	1.3 e ⁻⁴	2.45 e ⁻²
C6	<u>PRC2_EED_UP.V1_UP</u>	194	Genes up-regulated in TIG3 cells (fibroblasts) upon knockdown of EED [Gene ID=8726] gene.	3 (EVI5, DMD, SLC30A10)	0.015	3.05 e ⁻⁴	2.88 e ⁻²

Supplementary Table 19. Summary of impact of clonal and subclonal SNVs detected in DB17 tumour sample.

ANNOTATION	IMPACT	CLONAL	CLONAL	SUBCLONAL	SUBCLONAL	TOTAL	TOTAL	CONTROL	CONTROL
		CLL (N)	CLL (%)	CLL (N)	CLL (%)	CLL (N)	CLL (%)	T CELL (N)	T CELL (%)
stop gained	HIGH	1	0.03	0	0.00	1	0.03	1	0.01
sequence feature	MODERATE	3	0.08	0	0.00	3	0.08	1	0.01
missense variant & splice region variant	MODERATE	0	0.00	0	0.00	1	0.03	1	0.01
missense variant	MODERATE	10	0.27	0	0.00	10	0.26	1	0.01
intergenic region	MODIFIER	1660	45.42	49	29.70	1709	44.73	105	1.41
intron variant	MODIFIER	758	20.74	45	27.27	803	21.02	72	0.96
upstream gene variant	MODIFIER	286	7.82	13	7.88	299	7.83	46	0.62
downstream gene variant	MODIFIER	204	5.58	28	16.97	232	6.07	26	0.35
3' UTR variant	MODIFIER	24	0.66	0	0.00	24	0.63	2	0.03
non coding exon variant	MODIFIER	4	0.11	1	0.61	5	0.13	6	0.08
TF binding site	MODIFIER	1	0.03	0	0.00	1	0.03	0	0.00
5' UTR variant	MODIFIER	6	0.16	1	0.61	7	0.18	1	0.01
sequence feature	LOW	475	13.00	26	15.76	501	13.11	45	0.60
synonymous variant	LOW	3	0.08	2	1.21	5	0.13	0	0.00
5' UTR premature start codon gain variant	LOW	1	0.03	0	0.00	1	0.03	0	0.00
splice region variant	LOW	1	0.03	0	0.00	1	0.03	0	0.00
UNANNOTATED	UNKNOWN	4	0.11	0	0.00	4	0.10	2	0.03
not associated with disease	NONE (dbSNP)	214	5.85	0	0.00	214	5.60	7160	95.86
TOTAL		3655	100.0	165	100.0	3821	100.0	7469	100.00

*N= number of events

Supplementary Table 20. High and moderate impact mutations in tumour CLL-B cells (DB30).

Impact	Annotation	Position	R	A	Gene	Gene_ID	Prot_ID	Prot_Name	aa change	Impact*
HIGH	stop_gained	chr6:111695282	G	A	REV3L	ENSG00000009413	O60673	DNA polymerase zeta catalytic subunit	p.Gln1426*	-/-
MODERATE	missense_variant	chr12:7636230	G	A	CD163	ENSG00000177575	Q86VB7	Scavenger receptor cysteine-rich type 1 protein M130	p.Arg941Cys	1/1
MODERATE	missense_variant	chr16:3808027	T	G	CREBBP	ENSG00000005339	Q92793	CREB-binding protein	p.Asn1131Thr	0/1
MODERATE	missense_variant	chr17:7578395	G	A	TP53	ENSG00000141510	P04637	P04637	p.His179Tyr	1/1 †
MODERATE	missense_variant	chr7:82584442	T	C	PCLO	ENSG00000186472	Q9Y6V0	Protein piccolo	p.Lys1943Glu	0/0
MODERATE	missense_variant	chr7:90894846	G	T	FZD1	ENSG00000157240	Q9UP38	Frizzled-1	p.Lys217Asn	0/0
MODERATE	Missense.variant&splice.region.variant	chrX:114881991	A	G	PLS3	ENSG00000102024	P13797	Plastin-3	p.Lys545Glu	1/0
MODERATE	missense_variant	chr14:107170035	A	C	IGHV1-69	ENSG00000211973	P01742	Immunoglobulin heavy variable 1-69	p.Phe83Val	0/-
MODERATE	missense_variant	chrX:129264081	G	C	AIFM1	ENSG00000156709	O95831	Apoptosis-inducing factor 1, mitochondrial	p.Pro545Arg	1/0
MODERATE	missense_variant	chr12:44148838	G	T	PUS7L	ENSG00000129317	Q9H0K6	Pseudouridylate synthase 7 homolog-like protein	p.Pro71Thr	0/0 †
MODERATE	missense_variant	chr8:16043711	A	G	MSR1	ENSG00000038945	P21757	Macrophage scavenger receptor types I and II	p.Tyr3His	0/0
MODERATE	missense_variant	chr6:14118276	T	C	CD83	ENSG00000112149	Q01151	CD83 antigen	p.Tyr45His	1/1
MODERATE	sequence_feature	chr13:36885636	G	T	SPG20	ENSG00000133104	Q8N0X7	Spartin	-	-/-
MODERATE	sequence_feature	chr5:42644549	T	A	GHR	ENSG00000112964	P10912	Growth hormone receptor	-	-/-
MODERATE	sequence_feature	chr6:51683780	T	A	PKHD1	ENSG00000170927	P08F94	Fibrocystin	-	-/-

*Impact from SIFT and PROVEAN: not applicable/not found (-), neutral/tolerated (0), damaging/deleterious (1), not novel mutation (†).

Supplementary Table 21. List of genes with subclonal mutations.

Ensembl ID	Entrez ID	Gene Name
ENSG00000175054	545	ATR serine/threonine kinase(ATR)
ENSG00000265612	100616374	microRNA 4539(MIR4539)
ENSG00000074527	59277	netrin 4(NTN4)
ENSG00000183117	64478	CUB and Sushi multiple domains 1(CSMD1)
ENSG00000229656	101929475	uncharacterized LOC101929475(LOC101929475)
ENSG00000174469	26047	contactin associated protein-like 2(CNTNAP2)
ENSG00000122584	30010	neurexophilin 1(NXPH1)
ENSG00000170959	100506627	doublecortin domain containing 5(DCDC5)
ENSG00000240230	90639	COX19, cytochrome c oxidase assembly factor(COX19)
ENSG00000109452	8821	inositol polyphosphate-4-phosphatase type II B(INPP4B)
ENSG00000185274	64409	Williams-Beuren syndrome chromosome region 17(WBSCR17)
ENSG00000143995	4211	Meis homeobox 1(MEIS1)
ENSG00000253363	105379374	uncharacterized LOC105379374(LOC105379374)
ENSG00000136267	1607	diacylglycerol kinase beta(DGKB)
ENSG00000143951	51057	WD repeat containing planar cell polarity effector(WDPCP)
ENSG00000141905	4782	nuclear factor I C(NFIC)
ENSG00000110931	10645	calcium/calmodulin dependent protein kinase kinase 2(CAMKK2)
ENSG00000169439	6383	syndecan 2(SDC2)
ENSG00000169891	9185	RALBP1 associated Eps domain containing 2(REPS2)
ENSG00000151025	57512	G protein-coupled receptor 158(GPR158)
ENSG00000249464	285419	long intergenic non-protein coding RNA 1091(LINC01091)
ENSG00000196597	158431	zinc finger protein 782(ZNF782)
ENSG00000227906	100131208	SNAP25 antisense RNA 1(SNAP25-AS1)
ENSG00000004534	10180	RNA binding motif protein 6(RBM6)
ENSG00000171045	203062	t-SNARE domain containing 1(TSNARE1)
ENSG00000219481	55672	neuroblastoma breakpoint family member 1(NBPF1)
ENSG00000161551	84765	zinc finger protein 577(ZNF577)
ENSG00000261609	100616150	microRNA 4720(MIR4720)
ENSG00000152495	814	calcium/calmodulin dependent protein kinase IV(CAMK4)

ENSG00000150471	<u>23284</u>	adhesion G protein-coupled receptor L3(ADGRL3)
ENSG00000181090	<u>79813</u>	euchromatic histone lysine methyltransferase 1(EHMT1)
ENSG00000149972	<u>53942</u>	contactin 5(CNTN5)
ENSG00000142207	<u>9875</u>	URB1 ribosome biogenesis 1 homolog (<i>S. cerevisiae</i>)(URB1)
ENSG00000066855	<u>9650</u>	mitochondrial fission regulator 1(MTFR1)
ENSG00000260792	<u>101929634</u>	uncharacterized LOC101929634(LOC101929634)
ENSG00000154227	<u>204219</u>	ceramide synthase 3(CERS3)
ENSG00000165675	<u>10495</u>	ecto-NOX disulfide-thiol exchanger 2(ENOX2)
ENSG00000159079	<u>56683</u>	chromosome 21 open reading frame 59(C21orf59)
ENSG00000204174	<u>5540</u>	neuropeptide Y receptor Y4(NPY4R)
ENSG00000151322	<u>64067</u>	neuronal PAS domain protein 3(NPAS3)
ENSG00000165731	<u>5979</u>	ret proto-oncogene(RET)
ENSG00000003096	<u>90293</u>	kelch like family member 13(KLHL13)
ENSG00000013583	<u>50865</u>	heme binding protein 1(HEBP1)
ENSG00000241839	<u>80301</u>	pleckstrin homology domain containing O2(PLEKHO2)
ENSG00000164904	<u>501</u>	aldehyde dehydrogenase 7 family member A1(ALDH7A1)
ENSG00000219545	<u>729852</u>	UBAP1-MVB12-associated (UMA) domain containing 1(UMAD1)
ENSG00000241697	<u>8577</u>	transmembrane protein with EGF like and two follistatin like domains 1(TMEFF1)
ENSG00000157764	<u>673</u>	B-Raf proto-oncogene, serine/threonine kinase(BRAF)
ENSG00000168675	<u>753</u>	low density lipoprotein receptor class A domain containing 4(LDLRAD4)
ENSG00000168743	<u>255743</u>	nephronectin(NPNT)
ENSG00000261231	<u>101927132</u>	uncharacterized LOC101927132(LOC101927132)
ENSG00000139793	<u>10150</u>	muscleblind like splicing regulator 2(MBNL2)
ENSG00000176299	<u>441670</u>	olfactory receptor family 4 subfamily M member 1(OR4M1)
ENSG00000258779	<u>100506172</u>	long intergenic non-protein coding RNA 1568(LINC01568)
ENSG00000139174	<u>144165</u>	prickle planar cell polarity protein 1(PRICKLE1)
ENSG00000133106	<u>94240</u>	epithelial stromal interaction 1(EPSTI1)
ENSG00000144455	<u>285362</u>	sulfatase modifying factor 1(SUMF1)
ENSG00000171502	<u>255631</u>	collagen type XXIV alpha 1 chain(COL24A1)
ENSG00000183631	<u>100130613</u>	proline rich 32(PRR32)
ENSG00000221641	<u>100302233</u>	microRNA 1268a(MIR1268A)

ENSG00000198198	<u>23334</u>	seizure threshold 2 homolog (mouse)(SZT2)
ENSG00000140443	<u>3480</u>	insulin like growth factor 1 receptor(IGF1R)
ENSG00000065150	<u>3843</u>	importin 5(IPO5)
ENSG00000187391	<u>9863</u>	membrane associated guanylate kinase, WW and PDZ domain containing 2(MAGI2)
ENSG00000103489	<u>64131</u>	xylosyltransferase 1(XYLT1)
ENSG00000230487	<u>114796</u>	PSMG3 antisense RNA 1 (head to head)(PSMG3-AS1)
ENSG00000169933	<u>9758</u>	FERM and PDZ domain containing 4(FRMPD4)
ENSG00000104093	<u>23312</u>	Dmx like 2(DMXL2)
ENSG00000010438	<u>5646</u>	protease, serine 3(PRSS3)
ENSG00000137642	<u>6653</u>	sortilin related receptor 1(SORL1)
ENSG00000156650	<u>23522</u>	lysine acetyltransferase 6B(KAT6B)
ENSG00000007968	<u>1870</u>	E2F transcription factor 2(E2F2)
ENSG00000072133	<u>27330</u>	ribosomal protein S6 kinase A6(RPS6KA6)
ENSG00000165304	<u>9833</u>	maternal embryonic leucine zipper kinase(MELK)
ENSG00000107249	<u>169792</u>	GLIS family zinc finger 3(GLIS3)
ENSG00000198785	<u>116443</u>	glutamate ionotropic receptor NMDA type subunit 3A(GRIN3A)
ENSG00000162670	<u>339479</u>	BMP/retinoic acid inducible neural specific 3(BRINP3)
ENSG00000174738	<u>9975</u>	nuclear receptor subfamily 1 group D member 2(NR1D2)
ENSG00000135298	<u>577</u>	adhesion G protein-coupled receptor B3(ADGRB3)

Supplementary Table 22. Enrichment in gene ontology biological process from DB30 subclonal mutations.

DAVID assigned GO Term Biological Process	# genes within GO category	% genes within GO category	P-value	Benjamini-Hochberg corrected P-value
multicellular organismal process	34	43	2.70E-02	7.60E-01
single-multicellular organism process	30	38	2.60E-02	7.60E-01
single organism signalling	30	38	4.70E-02	8.10E-01
<i>ATR, BRAF, E2F2, GPR158, REPS2, WDPCP, ADGRB3, ADGRL3, CAMK4, CAMKK2, CNTNAP2, DGKB, DCDC5, EHMT1, GRIN3A, INPP4B, IGF1R, LDLRAD4, MELK, MAGI2, NPNT, NTN4, NPY4R, NR1D2, OR4M1, PRICKLE1, RET, RPS6KA6, SORL1, SDC2.</i>				
developmental process	29	36.7	2.40E-02	7.70E-01
anatomical structure development	28	35.4	3.00E-02	7.80E-01
single-organism developmental process	28	35.4	3.00E-02	7.70E-01

multicellular organism development	27	34.2	1.30E-02	7.50E-01
system development	24	30.4	2.00E-02	7.70E-01
cellular developmental process	23	29.1	1.70E-02	7.70E-01
cell differentiation	21	26.6	2.60E-02	7.60E-01
<i>BRAF, BRINP3, MEIS1, WDPCP, ADGRB3, ADGRL3, CAMK4, CERS3, COL24A1, CNTNAP2, GRIN3A, LDLRAD4, MAGI2, NPNT, NTN4, NR1D2, PRICKLE1, RET, SZT2, SORL1, SDC2.</i>				
regulation of multicellular organismal process	17	21.5	2.20E-02	7.70E-01
regulation of multicellular organismal development	16	20.3	8.40E-04	4.90E-01
regulation of developmental process	16	20.3	6.80E-03	7.90E-01
nervous system development	16	20.3	8.10E-03	7.30E-01
generation of neurons	14	17.7	8.40E-04	3.60E-01
neurogenesis	14	17.7	1.50E-03	4.60E-01
cell development	14	17.7	2.00E-02	7.50E-01
regulation of cell differentiation	13	16.5	7.60E-03	7.50E-01
neuron differentiation	12	15.2	3.70E-03	7.00E-01
positive regulation of macromolecule biosynthetic process	12	15.2	2.50E-02	7.60E-01
positive regulation of nucleobase-containing compound metabolic process	12	15.2	3.10E-02	7.80E-01
positive regulation of cellular biosynthetic process	12	15.2	4.10E-02	8.10E-01
<i>ATR, GLIS3, MEIS1, CAMK4, CAMKK2, IGF1R, KAT6B, NPNT, NPAS3, NR1D2, RET.</i>				
positive regulation of biosynthetic process	12	15.2	4.50E-02	8.20E-01
positive regulation of nitrogen compound metabolic process	12	15.2	4.70E-02	8.20E-01
cell projection organization	11	13.9	1.60E-02	7.70E-01
locomotion	11	13.9	4.30E-02	8.20E-01
neuron development	10	12.7	6.90E-03	7.50E-01
cell migration	10	12.7	2.40E-02	7.60E-01
<i>BRAF, WDPCP, ADGRL3, IGF1R, LDLRAD4, MAGI2, PRSS3, RET, SORL1, SDC2.</i>				
cell motility	10	12.7	4.50E-02	8.20E-01
localization of cell	10	12.7	4.50E-02	8.20E-01

positive regulation of nucleic acid-templated transcription	10	12.7	4.90E-02	8.10E-01
positive regulation of transcription, DNA-templated	10	12.7	4.90E-02	8.10E-01
<i>GLIS3, MEIS1, CAMK4, CAMKK2, KAT6B, NPNT, NPAS3, NFIC, NR1D2, RET.</i>				
regulation of nervous system development	9	11.4	5.80E-03	7.90E-01
neuron projection development	9	11.4	8.30E-03	7.00E-01
regulation of cell development	9	11.4	1.00E-02	7.30E-01
regulation of anatomical structure morphogenesis	9	11.4	3.30E-02	7.80E-01
regulation of neurogenesis	8	10.1	1.10E-02	7.00E-01
cell morphogenesis involved in differentiation	8	10.1	1.70E-02	7.40E-01
behaviour	7	8.9	1.40E-02	7.60E-01
regulation of neuron differentiation	7	8.9	1.50E-02	7.50E-01
regulation of cell migration	7	8.9	3.50E-02	7.90E-01
regulation of cell motility	7	8.9	4.70E-02	8.10E-01
positive regulation of nervous system development	6	7.6	2.20E-02	7.60E-01
morphogenesis of an epithelium	6	7.6	3.80E-02	8.00E-01
protein autophosphorylation	5	6.3	8.90E-03	7.00E-01
<i>ATR, CAMK4, CAMKK2, IGF1R, MELK.</i>				
regulation of cell morphogenesis involved in differentiation	5	6.3	3.30E-02	7.70E-01
neuron maturation	4	5.1	3.60E-04	4.40E-01
cell maturation	4	5.1	1.70E-02	7.60E-01
learning or memory	4	5.1	4.40E-02	8.20E-01
synapse organization	4	5.1	5.00E-02	8.10E-01
neuron remodelling	2	2.5	3.50E-02	7.80E-01

Supplementary Table 23. Clonal mutations are enriched in the MAPK cascade.

DAVID assigned GO Term Biological Process	# genes within GO category	% genes within GO category	P-value	Benjamini-Hochberg corrected P-value
MAPK cascade	52	5.5	5.30E-03	1.3E-01
ASH1 like histone lysine methyltransferase (<i>ASH1L</i>), B-cell scaffold protein with ankyrin repeats 1 (<i>BANK1</i>), BMP binding endothelial regulator (<i>BMPER</i>), CNKSR family member 3 (<i>CNKSR3</i>), EPH receptor A4 (<i>EPHA4</i>), EPH receptor A7 (<i>EPHA7</i>), LIM domain only 3 (<i>LMO3</i>), NIMA related kinase 10 (<i>NEK10</i>), RAS guanyl releasing protein 3 (<i>RASGRP3</i>), RAS protein activator like 2 (<i>RASAL2</i>), Ras like without CAAX 2 (<i>RIT2</i>), SAM and SH3 domain containing 1 (<i>SASH1</i>), SHC adaptor protein 3 (<i>SHC3</i>), TGF-beta activated kinase 1/MAP3K7 binding protein 3 (<i>TAB3</i>), TNF receptor associated factor 2 (<i>TRAF2</i>), TNF receptor superfamily member 11a (<i>TNFRSF11A</i>), TNFAIP3 interacting protein 1 (<i>TNIP1</i>), TRAF2 and NCK interacting kinase (<i>TNIK</i>), adrenoceptor alpha 1A (<i>ADRA1A</i>), angiopoietin 1 (<i>ANGPT1</i>), docking protein 5 (<i>DOK5</i>), dystrobrevin binding protein 1 (<i>DTNBP1</i>), erb-b2 receptor tyrosine kinase 4 (<i>ERBB4</i>), fibroblast growth factor 13 (<i>FGF13</i>), fibroblast growth factor 7 (<i>FGF7</i>), fibroblast growth factor receptor 2 (<i>FGFR2</i>), follicle stimulating hormone receptor (<i>FSHR</i>), forkhead box O1 (<i>FOXO1</i>), galectin 9 (<i>LGALS9</i>), glutamate ionotropic receptor NMDA type subunit 2A (<i>GRIN2A</i>), glutamate ionotropic receptor kainate type subunit 2 (<i>GRIK2</i>), growth hormone receptor (<i>GHR</i>), hepatocyte growth factor (<i>HGF</i>), insulin like growth factor 1 receptor (<i>IGF1R</i>), insulin receptor (<i>INSR</i>), interleukin 11 (<i>IL11</i>), mitogen-activated protein kinase 13 (<i>MAP3K13</i>), mitogen-activated protein kinase 7 (<i>MAP3K7</i>), neuregulin 1 (<i>NRG1</i>), neuregulin 4 (<i>NRG4</i>), parkin RBR E3 ubiquitin protein ligase (<i>PARK2</i>), phospholipase C beta 1 (<i>PLCB1</i>), platelet derived growth factor C (<i>PDGFC</i>), polycystic kidney and hepatic disease 1 (autosomal recessive) (<i>PKHD1</i>), proteasome 26S subunit, ATPase 6 (<i>PSMC6</i>), proteasome subunit beta 1 (<i>PSMB1</i>), protein kinase C alpha (<i>PRKCA</i>), protein tyrosine phosphatase, non-receptor type 2 (<i>PTPN2</i>), protein tyrosine phosphatase, receptor type J (<i>PTPRJ</i>), receptor tyrosine kinase like orphan receptor 2 (<i>ROR2</i>), spectrin alpha, erythrocytic 1 (<i>SPTA1</i>), sprouty related EVH1 domain containing 2 (<i>SPRED2</i>).				

Supplementary Table 24. 7 mutations disrupting POT1-TPP1 binding.

#	bp	aa	Location hg19	(+) strand	(-) strand	primer
1	G- >T/C	L343F	chr7:124482995	CTCTC <u>C</u> AAATA	TATTTG <u>G</u> GAGAG	POT3B
2	C->A	P446Q	chr7:124481059	AAAGC <u>G</u> GGAGA	TCTCC <u>G</u> GCTTT	POT3A
3	A->T	K465*	chr7:124475445	GAGTTT <u>G</u> CAAA	TTTGCA <u>A</u> ACTC	POT2B
4	C->T	P475L	chr7:124475414	TCACAG <u>G</u> AATT	AATTC <u>T</u> GTGA	POT2A
5	G->C	R477T	chr7:124475408	CAGAT <u>C</u> TCACA	TGTGAG <u>A</u> TCTG	POT2A
6	A->T	I535F	chr7:124467351	TACAAT <u>A</u> ACCCA	TGGGT <u>A</u> TTGTA	POT1B
7	T->G	C591W	chr7:124465325	GGAGG <u>A</u> CAAAA	TTTTG <u>T</u> CCTCC	POT1A

List of 7 mutations predicted to disrupt the POT1-TPP1 interaction. Mutation in the cDNA, change of aminoacid, chromosomal location and primers to sequence the mutation are indicated.

REFERENCES

- ALCARAZ SILVA, B., JONES, T. J. & MURNANE, J. P. 2017. Differences in the recruitment of DNA repair proteins at subtelomeric and interstitial I-SceI endonuclease-induced DNA double-strand breaks. *DNA Repair (Amst)*, 49, 1-8.
- ALEXANDROV, L. B., JONES, P. H., WEDGE, D. C., SALE, J. E., CAMPBELL, P. J., NIK-ZAINAL, S. & STRATTON, M. R. 2015. Clock-like mutational processes in human somatic cells. *Nat Genet*, 47, 1402-7.
- ALEXANDROV, L. B., NIK-ZAINAL, S., WEDGE, D. C., APARICIO, S. A., BEHJATI, S., BIANKIN, A. V., BIGNELL, G. R., BOLLI, N., BORG, A., BORRESEN-DALE, A. L., BOYAULT, S., BURKHARDT, B., BUTLER, A. P., CALDAS, C., DAVIES, H. R., DESMEDT, C., EILS, R., EYFJORD, J. E., FOEKENS, J. A., GREAVES, M., HOSODA, F., HUTTER, B., ILICIC, T., IMBEAUD, S., IMIELINSKI, M., JAGER, N., JONES, D. T., JONES, D., KNAPPSKOG, S., KOOL, M., LAKHANI, S. R., LOPEZ-OTIN, C., MARTIN, S., MUNSHI, N. C., NAKAMURA, H., NORTHCOTT, P. A., PAJIC, M., PAPAEMMANUIL, E., PARADISO, A., PEARSON, J. V., PUENTE, X. S., RAINE, K., RAMAKRISHNA, M., RICHARDSON, A. L., RICHTER, J., ROSENSTIEL, P., SCHLESNER, M., SCHUMACHER, T. N., SPAN, P. N., TEAGUE, J. W., TOTOKI, Y., TUTT, A. N., VALDES-MAS, R., VAN BUUREN, M. M., VAN 'T VEER, L., VINCENT-SALOMON, A., WADDELL, N., YATES, L. R., AUSTRALIAN PANCREATIC CANCER GENOME, I., CONSORTIUM, I. B. C., CONSORTIUM, I. M.-S., PEDBRAIN, I., ZUCMAN-ROSSI, J., FUTREAL, P. A., MCDERMOTT, U., LICHTER, P., MEYERSON, M., GRIMMOND, S. M., SIEBERT, R., CAMPO, E., SHIBATA, T., PFISTER, S. M., CAMPBELL, P. J. & STRATTON, M. R. 2013a. Signatures of mutational processes in human cancer. *Nature*, 500, 415-21.
- ALEXANDROV, L. B., NIK-ZAINAL, S., WEDGE, D. C., CAMPBELL, P. J. & STRATTON, M. R. 2013b. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep*, 3, 246-59.
- ALLSHIRE, R. C., DEMPSTER, M. & HASTIE, N. D. 1989. Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Res*, 17, 4611-27.
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol*, 215, 403-10.
- AMAYA-CHANAGA, C. I. & RASSENTI, L. Z. 2016. Biomarkers in chronic lymphocytic leukemia: Clinical applications and prognostic markers. *Best Pract Res Clin Haematol*, 29, 79-89.
- AMBROSINI, A., PAUL, S., HU, S. & RIETHMAN, H. 2007. Human subtelomeric duplicon structure and organization. *Genome Biol*, 8, R151.
- ANAND, R., RANJHA, L., CANNAVO, E. & CEJKA, P. 2016. Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol Cell*, 64, 940-950.
- ANCELIN, K., BRUNORI, M., BAUWENS, S., KOERING, C. E., BRUN, C., RICOUL, M., POMMIER, J. P., SABATIER, L. & GILSON, E. 2002. Targeting Assay To Study the cis Functions of Human Telomeric Proteins: Evidence for Inhibition of Telomerase by TRF1 and for Activation of Telomere Degradation by TRF2. *Molecular and Cellular Biology*, 22, 3474-3487.
- ANDOR, N., GRAHAM, T. A., JANSEN, M., XIA, L. C., AKTIPIS, C. A., PETRITSCH, C., JI, H. P. & MALEY, C. C. 2016. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med*, 22, 105-13.

- ARAKAWA, H., BEDNAR, T., WANG, M., PAUL, K., MLADENOV, E., BENCSIK-THEILEN, A. A. & ILIAKIS, G. 2012. Functional redundancy between DNA ligases I and III in DNA replication in vertebrate cells. *Nucleic Acids Res*, 40, 2599-610.
- ARAKAWA, H., HAUSCHILD, J. & BUERSTEDDE, J. M. 2002. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science*, 295, 1301-6.
- ARLT, M. F., DURKIN, S. G., RAGLAND, R. L. & GLOVER, T. W. 2006. Common fragile sites as targets for chromosome rearrangements. *DNA Repair (Amst)*, 5, 1126-35.
- ARNOULT, N. & KARLSEDER, J. 2015. Complex interactions between the DNA-damage response and mammalian telomeres. *Nat Struct Mol Biol*, 22, 859-66.
- ARTANDI, S. E. & ATTARDI, L. D. 2005. Pathways connecting telomeres and p53 in senescence, apoptosis, and cancer. *Biochem Biophys Res Commun*, 331, 881-90.
- ARTANDI, S. E., CHANG, S., LEE, S. L., ALSON, S., GOTTLIEB, G. J., CHIN, L. & DEPINHO, R. A. 2000. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature*, 406, 641-5.
- AUGEREAU, A., T'KINT DE ROODENBEKE, C., SIMONET, T., BAUWENS, S., HORARD, B., CALLANAN, M., LEROUX, D., JALLADES, L., SALLES, G., GILSON, E. & PONCET, D. 2011. Telomeric damage in early stage of chronic lymphocytic leukemia correlates with shelterin dysregulation. *Blood*, 118, 1316-22.
- AZZALIN, C. M. & LINGNER, J. 2015. Telomere functions grounding on TERRA firma. *Trends Cell Biol*, 25, 29-36.
- AZZALIN, C. M., REICHENBACH, P., KHORIAULI, L., GIULOTTO, E. & LINGNER, J. 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, 318, 798-801.
- BADIE, S., CARLOS, A. R., FOLIO, C., OKAMOTO, K., BOUWMAN, P., JONKERS, J. & TARSOUNAS, M. 2015. BRCA1 and CtIP promote alternative non-homologous end-joining at uncapped telomeres. *EMBO J*, 34, 410-24.
- BAIRD, D. M., BRITT-COMPTON, B., ROWSON, J., AMSO, N. N., GREGORY, L. & KIPLING, D. 2006. Telomere instability in the male germline. *Hum Mol Genet*, 15, 45-51.
- BAIRD, D. M., ROWSON, J., WYNFORD-THOMAS, D. & KIPLING, D. 2003. Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nat Genet*, 33, 203-7.
- BAIRD, D. M. A. J. J. A. N. J. R. 1995. Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human XpYp telomere *The EMBO Journal*, 14 5433-5443.
- BANDARIA, J. N., QIN, P., BERK, V., CHU, S. & YILDIZ, A. 2016. Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin. *Cell*, 164, 735-46.
- BARON, B. W., ANASTASI, J., BIES, J., REDDY, P. L., JOSEPH, L., THIRMAN, M. J., WROBLEWSKI, K., WOLFF, L. & BARON, J. M. 2014. GFI1B, EVI5, MYB--additional genes that cooperate with the human BCL6 gene to promote the development of lymphomas. *Blood Cells Mol Dis*, 52, 68-75.
- BARTON, O., NAUMANN, S. C., DIEMER-BIEHS, R., KUNZEL, J., STEINLAGE, M., CONRAD, S., MAKHARASHVILI, N., WANG, J., FENG, L., LOPEZ, B. S., PAULL, T. T., CHEN, J., JEGGO, P. A. & LOBRICH, M. 2014. Polo-like kinase 3 regulates CtIP during DNA double-strand break repair in G1. *J Cell Biol*, 206, 877-94.
- BASKAR, S., KWONG, K. Y., HOFER, T., LEVY, J. M., KENNEDY, M. G., LEE, E., STAUDT, L. M., WILSON, W. H., WIESTNER, A. & RADER, C. 2008. Unique cell surface expression of receptor tyrosine kinase ROR1 in human B-cell chronic lymphocytic leukemia. *Clin Cancer Res*, 14, 396-404.
- BASSAGANYAS, L., BEA, S., ESCARAMIS, G., TORNADOR, C., SALAVERRIA, I., ZAPATA, L., DREHSEL, O., FERREIRA, P. G., RODRIGUEZ-SANTIAGO, B., TUBIO, J. M., NAVARRO, A., MARTIN-GARCIA, D., LOPEZ, C., MARTINEZ-TRILLOS, A., LOPEZ-GUILLERMO, A.,

- GUT, M., OSSOWSKI, S., LOPEZ-OTIN, C., CAMPO, E. & ESTIVILL, X. 2013. Sporadic and reversible chromothripsis in chronic lymphocytic leukemia revealed by longitudinal genomic analysis. *Leukemia*, 27, 2376-9.
- BAUMANN, P. & CECH, T. R. 2001. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*, 292, 1171-5.
- BAUR, J. A., ZOU, Y., SHAY, J. W. & WRIGHT, W. E. 2001. Telomere position effect in human cells. *Science*, 292, 2075-7.
- BEBENEK, K., PEDERSEN, L. C. & KUNKEL, T. A. 2014. Structure-function studies of DNA polymerase lambda. *Biochemistry*, 53, 2781-92.
- BECHTER, O. E., EISTERER, W., PALL, G., HILBE, W., KUHR, T. & THALER, J. 1998. Telomere length and telomerase activity predict survival in patients with B cell chronic lymphocytic leukemia. *Cancer Res*, 58, 4918-22.
- BELANCIO, V. P., ROY-ENGEL, A. M. & DEININGER, P. L. 2010. All y'all need to know 'bout retroelements in cancer. *Semin Cancer Biol*, 20, 200-10.
- BELL, R. J., RUBE, H. T., KREIG, A., MANCINI, A., FOUSE, S. D., NAGARAJAN, R. P., CHOI, S., HONG, C., HE, D., PEKMEZCI, M., WIENCKE, J. K., WRENSCH, M. R., CHANG, S. M., WALSH, K. M., MYONG, S., SONG, J. S. & COSTELLO, J. F. 2015. Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science*, 348, 1036-9.
- BERNDT, S. I., CAMP, N. J., SKIBOLA, C. F., VIJAI, J., WANG, Z., GU, J., NIETERS, A., KELLY, R. S., SMEDBY, K. E., MONNEREAU, A., COZEN, W., COX, A., WANG, S. S., LAN, Q., TERAS, L. R., MACHADO, M., YEAGER, M., BROOKS-WILSON, A. R., HARTGE, P., PURDUE, M. P., BIRMANN, B. M., VAJDIC, C. M., COCCO, P., ZHANG, Y., GILES, G. G., ZELENIUCH-JACQUOTTE, A., LAWRENCE, C., MONTALVAN, R., BURDETT, L., HUTCHINSON, A., YE, Y., CALL, T. G., SHANAFELT, T. D., NOVAK, A. J., KAY, N. E., LIEBOW, M., CUNNINGHAM, J. M., ALLMER, C., HJALGRIM, H., ADAMI, H. O., MELBYE, M., GLIMELIUS, B., CHANG, E. T., GLENN, M., CURTIN, K., CANNON-ALBRIGHT, L. A., DIVER, W. R., LINK, B. K., WEINER, G. J., CONDE, L., BRACCI, P. M., RIBY, J., ARNETT, D. K., ZHI, D., LEACH, J. M., HOLLY, E. A., JACKSON, R. D., TINKER, L. F., BENAVENTE, Y., SALA, N., CASABONNE, D., BECKER, N., BOFFETTA, P., BRENNAN, P., FORETOVA, L., MAYNADIE, M., MCKAY, J., STAINES, A., CHAFFEE, K. G., ACHENBACH, S. J., VACHON, C. M., GOLDIN, L. R., STROM, S. S., LEIS, J. F., WEINBERG, J. B., CAPORASO, N. E., NORMAN, A. D., DE ROOS, A. J., MORTON, L. M., SEVERSON, R. K., RIBOLI, E., VINEIS, P., KAAKS, R., MASALA, G., WEIDERPASS, E., CHIRLAQUE, M. D., VERMEULEN, R. C., TRAVIS, R. C., SOUTHEY, M. C., MILNE, R. L., ALBANES, D., VIRTAMO, J., WEINSTEIN, S., CLAVEL, J., ZHENG, T., HOLFORD, T. R., VILLANO, D. J., MARIA, A., SPINELLI, J. J., GASCOYNE, R. D., *et al.* 2016. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat Commun*, 7, 10933.
- BERNDT, S. I., SKIBOLA, C. F., JOSEPH, V., CAMP, N. J., NIETERS, A., WANG, Z., COZEN, W., MONNEREAU, A., WANG, S. S., KELLY, R. S., LAN, Q., TERAS, L. R., CHATTERJEE, N., CHUNG, C. C., YEAGER, M., BROOKS-WILSON, A. R., HARTGE, P., PURDUE, M. P., BIRMANN, B. M., ARMSTRONG, B. K., COCCO, P., ZHANG, Y., SEVERI, G., ZELENIUCH-JACQUOTTE, A., LAWRENCE, C., BURDETTE, L., YUENGER, J., HUTCHINSON, A., JACOBS, K. B., CALL, T. G., SHANAFELT, T. D., NOVAK, A. J., KAY, N. E., LIEBOW, M., WANG, A. H., SMEDBY, K. E., ADAMI, H. O., MELBYE, M., GLIMELIUS, B., CHANG, E. T., GLENN, M., CURTIN, K., CANNON-ALBRIGHT, L. A., JONES, B., DIVER, W. R., LINK, B. K., WEINER, G. J., CONDE, L., BRACCI, P. M., RIBY, J., HOLLY, E. A., SMITH, M. T., JACKSON, R. D., TINKER, L. F., BENAVENTE, Y., BECKER, N., BOFFETTA, P., BRENNAN, P., FORETOVA, L., MAYNADIE, M., MCKAY, J., STAINES, A., RABE, K. G., ACHENBACH, S. J., VACHON, C. M., GOLDIN, L. R., STROM,

- S. S., LANASA, M. C., SPECTOR, L. G., LEIS, J. F., CUNNINGHAM, J. M., WEINBERG, J. B., MORRISON, V. A., CAPORASO, N. E., NORMAN, A. D., LINET, M. S., DE ROOS, A. J., MORTON, L. M., SEVERSON, R. K., RIBOLI, E., VINEIS, P., KAAKS, R., TRICHOPOULOS, D., MASALA, G., WEIDERPASS, E., CHIRLAQUE, M. D., VERMEULEN, R. C., TRAVIS, R. C., GILES, G. G., ALBANES, D., VIRTAMO, J., WEINSTEIN, S., CLAVEL, J., ZHENG, T., HOLFORD, T. R., OFFIT, K., ZELENETZ, A., KLEIN, R. J., SPINELLI, J. J., BERTRAND, K. A., *et al.* 2013. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat Genet*, 45, 868-76.
- BEROUKHIM, R., MERMEL, C. H., PORTER, D., WEI, G., RAYCHAUDHURI, S., DONOVAN, J., BARRETINA, J., BOEHM, J. S., DOBSON, J., URASHIMA, M., MC HENRY, K. T., PINCHBACK, R. M., LIGON, A. H., CHO, Y. J., HAERY, L., GREULICH, H., REICH, M., WINCKLER, W., LAWRENCE, M. S., WEIR, B. A., TANAKA, K. E., CHIANG, D. Y., BASS, A. J., LOO, A., HOFFMAN, C., PRENSNER, J., LIEFELD, T., GAO, Q., YECIES, D., SIGNORETTI, S., MAHER, E., KAYE, F. J., SASAKI, H., TEPPER, J. E., FLETCHER, J. A., TABERNERO, J., BASELGA, J., TSAO, M. S., DEMICHELIS, F., RUBIN, M. A., JANNE, P. A., DALY, M. J., NUCERA, C., LEVINE, R. L., EBERT, B. L., GABRIEL, S., RUSTGI, A. K., ANTONESCU, C. R., LADANYI, M., LETAI, A., GARRAWAY, L. A., LODA, M., BEER, D. G., TRUE, L. D., OKAMOTO, A., POMEROY, S. L., SINGER, S., GOLUB, T. R., LANDER, E. S., GETZ, G., SELLERS, W. R. & MEYERSON, M. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature*, 463, 899-905.
- BHARGAVA, R., ONYANGO, D. O. & STARK, J. M. 2016. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet*, 32, 566-75.
- BHAT, A., ANDERSEN, P. L., QIN, Z. & XIAO, W. 2013. Rev3, the catalytic subunit of Polzeta, is required for maintaining fragile site stability in human cells. *Nucleic Acids Res*, 41, 2328-39.
- BIEHS, R., STEINLAGE, M., BARTON, O., JUHASZ, S., KUNZEL, J., SPIES, J., SHIBATA, A., JEGGO, P. A. & LOBRICH, M. 2017. DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination. *Mol Cell*, 65, 671-684 e5.
- BIENVENU, F., JIRAWATNOTAI, S., ELIAS, J. E., MEYER, C. A., MIZERACKA, K., MARSON, A., FRAMPTON, G. M., COLE, M. F., ODOM, D. T., ODAJIMA, J., GENG, Y., ZAGOZDZON, A., JECROIS, M., YOUNG, R. A., LIU, X. S., CEPKO, C. L., GYGI, S. P. & SICINSKI, P. 2010. Transcriptional role of cyclin D1 in development revealed by a genetic-proteomic screen. *Nature*, 463, 374-8.
- BIFFI, G., TANNAHILL, D., MCCAFFERTY, J. & BALASUBRAMANIAN, S. 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem*, 5, 182-6.
- BILBAN, M., HEINTEL, D., SCHARL, T., WOELFEL, T., AUER, M. M., PORPACZY, E., KAINZ, B., KROBER, A., CAREY, V. J., SHEHATA, M., ZIELINSKI, C., PICKL, W., STILGENBAUER, S., GAIGER, A., WAGNER, O., JAGER, U. & GERMAN, C. L. L. S. G. 2006. Deregulated expression of fat and muscle genes in B-cell chronic lymphocytic leukemia with high lipoprotein lipase expression. *Leukemia*, 20, 1080-8.
- BINET, J. L., LEPORRIER, M., DIGHIRO, G., CHARRON, D., D'ATHIS, P., VAUGIER, G., BERAL, H. M., NATALI, J. C., RAPHAEL, M., NIZET, B. & FOLLEZOU, J. Y. 1977. A clinical staging system for chronic lymphocytic leukemia: prognostic significance. *Cancer*, 40, 855-64.
- BIRKBAK, N. J., EKLUND, A. C., LI, Q., MCCLELLAND, S. E., ENDESFELDER, D., TAN, P., TAN, I. B., RICHARDSON, A. L., SZALLASI, Z. & SWANTON, C. 2011. Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res*, 71, 3447-52.
- BLACKBURN, E. H. & COLLINS, K. 2011. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol*, 3.

- BLACKBURN, E. H. & GALL, J. G. 1978. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J Mol Biol*, 120, 33-53.
- BODNAR, A. G., OUELLETTE, M., FROLKIS, M., HOLT, S. E., CHIU, C. P., MORIN, G. B., HARLEY, C. B., SHAY, J. W., LICHTSTEINER, S. & WRIGHT, W. E. 1998. Extension of life-span by introduction of telomerase into normal human cells. *Science*, 279, 349-52.
- BOERSMA, V., MOATTI, N., SEGURA-BAYONA, S., PEUSCHER, M. H., VAN DER TORRE, J., WEVERS, B. A., ORTHWEIN, A., DUROCHER, D. & JACOBS, J. J. L. 2015. MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5' end resection. *Nature*, 521, 537-540.
- BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114-20.
- BOSCO, N. & DE LANGE, T. 2012. A TRF1-controlled common fragile site containing interstitial telomeric sequences. *Chromosoma*, 121, 465-74.
- BOULIANNE, B., ROBINSON, M. E., MAY, P. C., CASTELLANO, L., BLIGHE, K., THOMAS, J., REID, A., MUSCHEN, M., APPERLEY, J. F., STEBBING, J. & FELDHAHN, N. 2017. Lineage-Specific Genes Are Prominent DNA Damage Hotspots during Leukemic Transformation of B Cell Precursors. *Cell Rep*, 18, 1687-1698.
- 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. *Hum Mol Genet*, 15, 725-33.
- BROCCOLI, D., CHONG, L., OELMANN, S., FERNALD, A. A., MARZILIANO, N., VAN STEENSEL, B., KIPLING, D., LE BEAU, M. M. & DE LANGE, T. 1997a. Comparison of the human and mouse genes encoding the telomeric protein, TRF1: chromosomal localization, expression and conserved protein domains. *Hum Mol Genet*, 6, 69-76.
- BROCCOLI, D., SMOGORZEWSKA, A., CHONG, L. & DE LANGE, T. 1997b. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat Genet*, 17, 231-5.
- BROWN, E. J. & BALTIMORE, D. 2003. Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. *Genes Dev*, 17, 615-28.
- BROWN, J. R., BYRD, J. C., COUTRE, S. E., BENSON, D. M., FLINN, I. W., WAGNER-JOHNSTON, N. D., SPURGEON, S. E., KAHL, B. S., BELLO, C., WEBB, H. K., JOHNSON, D. M., PETERMAN, S., LI, D., JAHN, T. M., LANNUTTI, B. J., ULRICH, R. G., YU, A. S., MILLER, L. L. & FURMAN, R. R. 2014. Idelalisib, an inhibitor of phosphatidylinositol 3-kinase p110delta, for relapsed/refractory chronic lymphocytic leukemia. *Blood*, 123, 3390-7.
- BROWN, W. R., MACKINNON, P. J., VILLASANTE, A., SPURR, N., BUCKLE, V. J. & DOBSON, M. J. 1990. Structure and polymorphism of human telomere-associated DNA. *Cell*, 63, 119-32.
- BURNS, A., ALSOLAMI, R., BECQ, J., TIMBS, A., BRUCE, D., ROBBE, P., VAVOULIS, D., CABES, M., DREAU, H., TAYLOR, J., KNIGHT, S. J. L., MANSSON, R., BENTLEY, D., BEEKMAN, R., MARTIN-SUBERO, J. I., CAMPO, E., HOULSTON, R. S., RIDOUT, K. E. & SCHUH, A. 2017. Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHVmut and IgHVunmut subgroups. *Leukemia*.
- BURROW, A. A., WILLIAMS, L. E., PIERCE, L. C. & WANG, Y. H. 2009. Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites. *BMC Genomics*, 10, 59.
- BYRD, J. C., FURMAN, R. R., COUTRE, S. E., BURGER, J. A., BLUM, K. A., COLEMAN, M., WIERDA, W. G., JONES, J. A., ZHAO, W., HEEREMA, N. A., JOHNSON, A. J., SHAW, Y., BILOTTI, E., ZHOU, C., JAMES, D. F. & O'BRIEN, S. 2015. Three-year follow-up of

- treatment-naive and previously treated patients with CLL and SLL receiving single-agent ibrutinib. *Blood*, 125, 2497-506.
- CAHILL, D. P., KINZLER, K. W., VOGELSTEIN, B. & LENGAUER, C. 1999. Genetic instability and darwinian selection in tumours. *Trends Cell Biol*, 9, M57-60.
- CAPPER, R., BRITT-COMPTON, B., TANKIMANOVA, M., ROWSON, J., LETSOLO, B., MAN, S., HAUGHTON, M. & BAIRD, D. M. 2007. The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes Dev*, 21, 2495-508.
- CAREW, J. S., NAWROCKI, S. T., XU, R. H., DUNNER, K., MCCONKEY, D. J., WIERDA, W. G., KEATING, M. J. & HUANG, P. 2004. Increased mitochondrial biogenesis in primary leukemia cells: the role of endogenous nitric oxide and impact on sensitivity to fludarabine. *Leukemia*, 18, 1934-40.
- CASELLAS, R., BASU, U., YEWDELL, W. T., CHAUDHURI, J., ROBBIANI, D. F. & DI NOIA, J. M. 2016. Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nat Rev Immunol*, 16, 164-76.
- CAWTHON, R. M. 2002. Telomere measurement by quantitative PCR. *Nucleic Acids Res*, 30, e47.
- CAWTHON, R. M. 2009. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res*, 37, e21.
- CAYROU, C., BALLESTER, B., PEIFFER, I., FENOUIL, R., COULOMBE, P., ANDRAU, J. C., VAN HELDEN, J. & MECHALI, M. 2015. The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res*, 25, 1873-85.
- CECCALDI, R., LIU, J. C., AMUNUGAMA, R., HAJDU, I., PRIMACK, B., PETALCORIN, M. I., O'CONNOR, K. W., KONSTANTINOPOULOS, P. A., ELLEDGE, S. J., BOULTON, S. J., YUSUFZAI, T. & D'ANDREA, A. D. 2015. Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. *Nature*, 518, 258-62.
- CECCALDI, R., RONDINELLI, B. & D'ANDREA, A. D. 2016. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol*, 26, 52-64.
- CELLI, G. B. & DE LANGE, T. 2005. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nat Cell Biol*, 7, 712-8.
- CERHAN, J. R. & SLAGER, S. L. 2015. Familial predisposition and genetic risk factors for lymphoma. *Blood*, 126, 2265-73.
- CESARE, A. J. & GRIFFITH, J. D. 2004. Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops. *Mol Cell Biol*, 24, 9948-57.
- CHAE, H. D., MITTON, B., LACAYO, N. J. & SAKAMOTO, K. M. 2015. Replication factor C3 is a CREB target gene that regulates cell cycle progression through the modulation of chromatin loading of PCNA. *Leukemia*, 29, 1379-89.
- CHAI, W., DU, Q., SHAY, J. W. & WRIGHT, W. E. 2006. Human telomeres have different overhang sizes at leading versus lagging strands. *Mol Cell*, 21, 427-35.
- CHAN, S. W. & BLACKBURN, E. H. 2003. Telomerase and ATM/Tel1p protect telomeres from nonhomologous end joining. *Mol Cell*, 11, 1379-87.
- CHANG, H. H., WATANABE, G., GERODIMOS, C. A., OCHI, T., BLUNDELL, T. L., JACKSON, S. P. & LIEBER, M. R. 2016. Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. *J Biol Chem*, 291, 24377-24389.
- CHANG, H. H., WATANABE, G. & LIEBER, M. R. 2015. Unifying the DNA end-processing roles of the artemis nuclease: Ku-dependent artemis resection at blunt DNA ends. *J Biol Chem*, 290, 24036-50.
- CHANG, H. H. Y., PANNUNZIO, N. R., ADACHI, N. & LIEBER, M. R. 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol*, 18, 495-506.

- CHEN, L. L. & YANG, L. 2017. ALU alternative Regulation for Gene Expression. *Trends Cell Biol*, 27, 480-490.
- CHENG, Q., BARBOULE, N., FRIT, P., GOMEZ, D., BOMBARDE, O., COUDERC, B., REN, G. S., SALLES, B. & CALSOU, P. 2011. Ku counteracts mobilization of PARP1 and MRN in chromatin damaged with DNA double-strand breaks. *Nucleic Acids Res*, 39, 9605-19.
- CHIBA, K., LORBEER, F. K., SHAIN, A. H., MCSWIGGEN, D. T., SCHRUF, E., OH, A., RYU, J., DARZACQ, X., BASTIAN, B. C. & HOCKEMEYER, D. 2017. Mutations in the promoter of the telomerase gene TERT contribute to tumorigenesis by a two-step mechanism. *Science*.
- CHO, N. W., DILLEY, R. L., LAMPSON, M. A. & GREENBERG, R. A. 2014. Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell*, 159, 108-21.
- CHOI, J. H., LINDSEY-BOLTZ, L. A., KEMP, M., MASON, A. C., WOLD, M. S. & SANCAR, A. 2010. Reconstitution of RPA-covered single-stranded DNA-activated ATR-Chk1 signaling. *Proc Natl Acad Sci U S A*, 107, 13660-5.
- CHOI, Y. & CHAN, A. P. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31, 2745-7.
- CHOI, Y., SIMS, G. E., MURPHY, S., MILLER, J. R. & CHAN, A. P. 2012. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*, 7, e46688.
- CIBULSKIS, K., LAWRENCE, M. S., CARTER, S. L., SIVACHENKO, A., JAFFE, D., SOUGNEZ, C., GABRIEL, S., MEYERSON, M., LANDER, E. S. & GETZ, G. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*, 31, 213-9.
- CINGOLANI, P., PLATTS, A., WANG LE, L., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. & RUDEN, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, 6, 80-92.
- COLEMAN, J., BAIRD, D. M. & ROYLE, N. J. 1999. The plasticity of human telomeres demonstrated by a hypervariable telomere repeat array that is located on some copies of 16p and 16q. *Hum Mol Genet*, 8, 1637-46.
- COLGIN, L. M., BARAN, K., BAUMANN, P., CECH, T. R. & REDDEL, R. R. 2003. Human POT1 Facilitates Telomere Elongation by Telomerase. *Current Biology*, 13, 942-946.
- 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. *J Cell Biol*, 199, 893-906.
- COOK, G. P., TOMLINSON, I. M., WALTER, G., RIETHMAN, H., CARTER, N. P., BULUWELA, L., WINTER, G. & RABBITTS, T. H. 1994. A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. *Nat Genet*, 7, 162-8.
- COQUELLE, A., PIPIRAS, E., TOLEDO, F., BUTTIN, G. & DEBATISSE, M. 1997. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell*, 89, 215-25.
- COUNTER, C. M., AVILION, A. A., LEFEUVRE, C. E., STEWART, N. G., GREIDER, C. W., HARLEY, C. B. & BACCHETTI, S. 1992. Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J*, 11, 1921-9.
- CROSSEN, P. E., TULLY, S. M., BENJES, S. M., HOLLINGS, P. E., BEARD, M. E., NIMMO, J. C. & MORRISON, M. J. 1993. Oligoclonal B-cell leukemia characterized by spontaneous cell division and telomere association. *Genes Chromosomes Cancer*, 8, 49-59.

- CUNEO, A., RIGOLIN, G. M., BIGONI, R., DE ANGELI, C., VERONESE, A., CAVAZZINI, F., BARDI, A., ROBERTI, M. G., TAMMISO, E., AGOSTINI, P., CICCONE, M., DELLA PORTA, M., TIEGHI, A., CAVAZZINI, L., NEGRINI, M. & CASTOLDI, G. 2004. Chronic lymphocytic leukemia with 6q- shows distinct hematological features and intermediate prognosis. *Leukemia*, 18, 476-83.
- CUSANELLI, E. & CHARTRAND, P. 2015. Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. *Front Genet*, 6, 143.
- DAI, X., HUANG, C., BHUSARI, A., SAMPATHI, S., SCHUBERT, K. & CHAI, W. 2010. Molecular steps of G-overhang generation at human telomeres and its function in chromosome end protection. *EMBO J*, 29, 2788-801.
- DAMLE, R. N., WASIL, T., FAIS, F., GHIOTTO, F., VALETTO, A., ALLEN, S. L., BUCHBINDER, A., BUDMAN, D., DITTMAR, K., KOLITZ, J., LICHTMAN, S. M., SCHULMAN, P., VINCIGUERRA, V. P., RAI, K. R., FERRARINI, M. & CHIORAZZI, N. 1999. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*, 94, 1840-7.
- DANESHMANESH, A. H., HOJJAT-FARSANGI, M., MOSHFEGH, A., KHAN, A. S., MIKAELSSON, E., OSTERBORG, A. & MELLSTEDT, H. 2015. The PI3K/AKT/mTOR pathway is involved in direct apoptosis of CLL cells induced by ROR1 monoclonal antibodies. *Br J Haematol*, 169, 455-8.
- DANESHMANESH, A. H., MIKAELSSON, E., JEDDI-TEHRANI, M., BAYAT, A. A., GHODS, R., OSTADKARAMPOUR, M., AKHONDI, M., LAGERCRANTZ, S., LARSSON, C., OSTERBORG, A., SHOKRI, F., MELLSTEDT, H. & RABBANI, H. 2008. Ror1, a cell surface receptor tyrosine kinase is expressed in chronic lymphocytic leukemia and may serve as a putative target for therapy. *Int J Cancer*, 123, 1190-5.
- DAVOLI, T. & DE LANGE, T. 2012. Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse cells. *Cancer Cell*, 21, 765-76.
- DAVOLI, T., DENCHI, E. L. & DE LANGE, T. 2010. Persistent telomere damage induces bypass of mitosis and tetraploidy. *Cell*, 141, 81-93.
- DAWSON, M. A. & KOUZARIDES, T. 2012. Cancer epigenetics: from mechanism to therapy. *Cell*, 150, 12-27.
- DAYAMA, G., EMERY, S. B., KIDD, J. M. & MILLS, R. E. 2014. The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res*, 42, 12640-9.
- DE LANGE, T. 2005a. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev*, 19, 2100-10.
- DE LANGE, T. 2005b. Telomere-related genome instability in cancer. *Cold Spring Harb Symp Quant Biol*, 70, 197-204.
- DE LANGE, T., SHIUE, L., MYERS, R. M., COX, D. R., NAYLOR, S. L., KILLERY, A. M. & VARMUS, H. E. 1990. Structure and variability of human chromosome ends. *Mol Cell Biol*, 10, 518-27.
- DEL GIUDICE, I., MARINELLI, M., WANG, J., BONINA, S., MESSINA, M., CHIARETTI, S., ILARI, C., CAFFORIO, L., RAPONI, S., MAURO, F. R., DI MAIO, V., DE PROPRIIS, M. S., NANNI, M., CIARDULLO, C., ROSSI, D., GAIDANO, G., GUARINI, A., RABADAN, R. & FOA, R. 2016. Inter- and intra-patient clonal and subclonal heterogeneity of chronic lymphocytic leukaemia: evidences from circulating and lymph nodal compartments. *Br J Haematol*, 172, 371-383.
- DELGADO, J., VILLAMOR, N., LOPEZ-GUILLERMO, A. & CAMPO, E. 2016. Genetic evolution in chronic lymphocytic leukaemia. *Best Pract Res Clin Haematol*, 29, 67-78.

- DEMIR, H. A., BAYHAN, T., UNER, A., KURTULAN, O., KARAKUS, E., EMIR, S., OZYORUK, D. & CEYLANER, S. 2014. Chronic lymphocytic leukemia in a child: a challenging diagnosis in pediatric oncology practice. *Pediatr Blood Cancer*, 61, 933-5.
- DENCHI, E. L. & DE LANGE, T. 2007. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature*, 448, 1068-71.
- DENG, W. & LUCAS, J. N. 1999. Combined FISH with pan-telomeric PNA and whole chromosome-specific DNA probes to detect complete and incomplete chromosomal exchanges in human lymphocytes. *Int J Radiat Biol*, 75, 1107-12.
- DENG, W., TSAO, S. W., GUAN, X. Y., LUCAS, J. N., SI, H. X., LEUNG, C. S., MAK, P., WANG, L. D. & CHEUNG, A. L. 2004. Distinct profiles of critically short telomeres are a key determinant of different chromosome aberrations in immortalized human cells: whole-genome evidence from multiple cell lines. *Oncogene*, 23, 9090-101.
- DENG, Y., CHAN, S. S. & CHANG, S. 2008. Telomere dysfunction and tumour suppression: the senescence connection. *Nat Rev Cancer*, 8, 450-8.
- DENG, Z., WANG, Z., STONG, N., PLASSCHAERT, R., MOCZAN, A., CHEN, H. S., HU, S., WIKRAMASINGHE, P., DAVULURI, R. V., BARTOLOMEI, M. S., RIETHMAN, H. & LIEBERMAN, P. M. 2012. A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J*, 31, 4165-78.
- DEPRISTO, M. A., BANKS, E., POPLIN, R., GARIMELLA, K. V., MAGUIRE, J. R., HARTL, C., PHILIPPAKIS, A. A., DEL ANGEL, G., RIVAS, M. A., HANNA, M., MCKENNA, A., FENNEL, T. J., KERNYTSKY, A. M., SIVACHENKO, A. Y., CIBULSKIS, K., GABRIEL, S. B., ALTSHULER, D. & DALY, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, 43, 491-8.
- DER-SARKISSIAN, H., BACCHETTI, S., CAZES, L. & LONDONO-VALLEJO, J. A. 2004. The shortest telomeres drive karyotype evolution in transformed cells. *Oncogene*, 23, 1221-8.
- DHILLON, A. S., HAGAN, S., RATH, O. & KOLCH, W. 2007. MAP kinase signalling pathways in cancer. *Oncogene*, 26, 3279-90.
- DILLEY, R. L. & GREENBERG, R. A. 2015. ALternative Telomere Maintenance and Cancer. *Trends Cancer*, 1, 145-156.
- DIXON, Z. A., NICHOLSON, L., ZEPPEZAUER, M., MATHESON, E., SINCLAIR, P., HARRISON, C. J. & IRVING, J. A. 2017. CREBBP knockdown enhances RAS/RAF/MEK/ERK signaling in Ras pathway mutated acute lymphoblastic leukemia but does not modulate chemotherapeutic response. *Haematologica*, 102, 736-745.
- DOHNER, H., STILGENBAUER, S., BENNER, A., LEUPOLT, E., KROBER, A., BULLINGER, L., DOHNER, K., BENTZ, M. & LICHTER, P. 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*, 343, 1910-6.
- DOHNER, H., STILGENBAUER, S., JAMES, M. R., BENNER, A., WEILGUNI, T., BENTZ, M., FISCHER, K., HUNSTEIN, W. & LICHTER, P. 1997. 11q deletions identify a new subset of B-cell chronic lymphocytic leukemia characterized by extensive nodal involvement and inferior prognosis. *Blood*, 89, 2516-22.
- DOKSANI, Y. & DE LANGE, T. 2016. Telomere-Internal Double-Strand Breaks Are Repaired by Homologous Recombination and PARP1/Lig3-Dependent End-Joining. *Cell Rep*, 17, 1646-1656.
- DONG, J. T. 2001. Chromosomal deletions and tumor suppressor genes in prostate cancer. *Cancer Metastasis Rev*, 20, 173-93.
- DUCRAY, C., POMMIER, J. P., MARTINS, L., BOUSSIN, F. D. & SABATIER, L. 1999. Telomere dynamics, end-to-end fusions and telomerase activation during the human fibroblast immortalization process. *Oncogene*, 18, 4211-23.

- DUNHAM, M. A., NEUMANN, A. A., FASCHING, C. L. & REDDEL, R. R. 2000. Telomere maintenance by recombination in human cells. *Nat Genet*, 26, 447-50.
- DURIG, J., NASCHAR, M., SCHMUCKER, U., RENZING-KOHLER, K., HOLTER, T., HUTTMANN, A. & DUHRSEN, U. 2002. CD38 expression is an important prognostic marker in chronic lymphocytic leukaemia. *Leukemia*, 16, 30-5.
- DURIG, J., NUCKEL, H., CREMER, M., FUHRER, A., HALFMEYER, K., FANDREY, J., MOROY, T., KLEIN-HITPASS, L. & DUHRSEN, U. 2003. ZAP-70 expression is a prognostic factor in chronic lymphocytic leukemia. *Leukemia*, 17, 2426-34.
- DWYER, J. M. & LIU, J. P. 2010. Ets2 transcription factor, telomerase activity and breast cancer. *Clin Exp Pharmacol Physiol*, 37, 83-7.
- DZHAGALOV, I., GIGUÈRE, V. & HE, Y.-W. 2004. Lymphocyte Development and Function in the Absence of Retinoic Acid-Related Orphan Receptor α . *The Journal of Immunology*, 173, 2952-2959.
- EDELMANN, J., HOLZMANN, K., MILLER, F., WINKLER, D., BUHLER, A., ZENZ, T., BULLINGER, L., KUHN, M. W., GERHARDINGER, A., BLOEHDORN, J., RADTKE, I., SU, X., MA, J., POUNDS, S., HALLEK, M., LICHTER, P., KORBEL, J., BUSCH, R., MERTENS, D., DOWNING, J. R., STILGENBAUER, S. & DOHNER, H. 2012. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood*, 120, 4783-94.
- EDLING, A. E., NANAVATI, T., JOHNSON, J. M. & TUOHY, V. K. 2004. Human and murine lymphocyte neurotrophin expression is confined to B cells. *J Neurosci Res*, 77, 709-17.
- EGAN, E. D. & COLLINS, K. 2010. Specificity and stoichiometry of subunit interactions in the human telomerase holoenzyme assembled *in vivo*. *Mol Cell Biol*, 30, 2775-86.
- ELLIOTT, B., RICHARDSON, C. & JASIN, M. 2005. Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. *Mol Cell*, 17, 885-94.
- ERNST, A., JONES, D. T., MAASS, K. K., RODE, A., DEEG, K. I., JEBARAJ, B. M., KORSHUNOV, A., HOVESTADT, V., TAINSKY, M. A., PAJTLER, K. W., BENDER, S., BRABETZ, S., GROBNER, S., KOOL, M., DEVENS, F., EDELMANN, J., ZHANG, C., CASTELO-BRANCO, P., TABORI, U., MALKIN, D., RIPPE, K., STILGENBAUER, S., PFISTER, S. M., ZAPATKA, M. & LICHTER, P. 2016. Telomere dysfunction and chromothripsis. *Int J Cancer*, 138, 2905-14.
- ESCRIBANO-DIAZ, C., ORTHWEIN, A., FRADET-TURCOTTE, A., XING, M., YOUNG, J. T., TKAC, J., COOK, M. A., ROSEBROCK, A. P., MUNRO, M., CANNY, M. D., XU, D. & DUROCHER, D. 2013. A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol Cell*, 49, 872-83.
- ETXABE, A., LARA-CASTILLO, M. C., CORNET-MASANA, J. M., BANUS-MULET, A., NOMDEDEU, M., TORRENTE, M. A., PRATCORONA, M., DIAZ-BEYA, M., ESTEVE, J. & RISUENO, R. M. 2017. Inhibition of serotonin receptor type 1 in acute myeloid leukemia impairs leukemia stem cell functionality: a promising novel therapeutic target. *Leukemia*.
- FABRIZIO D'ADDA DI FAGAGNA, P. M. R., LORENA CLAY-FARRACE,, HEIKE FIEGLER, P. C., THOMAS VON ZGLINICKI, GABRIELE SARETZKI, & JACKSON, N. P. C. S. P. 2003. A dna damage checkpoint response in telomere-initiated senescence. *Nature*, 426.
- FEINBERG, A. P., KOLDOBSKIY, M. A. & GONDOR, A. 2016. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet*, 17, 284-99.
- FERREIRA, P. G., JARES, P., RICO, D., GOMEZ-LOPEZ, G., MARTINEZ-TRILLOS, A., VILLAMOR, N., ECKER, S., GONZALEZ-PEREZ, A., KNOWLES, D. G., MONLONG, J., JOHNSON, R., QUESADA, V., DJEBALI, S., PAPASAIKAS, P., LOPEZ-GUERRA, M., COLOMER, D., ROYO, C., CAZORLA, M., PINYOL, M., CLOT, G., AYMERICH, M., ROZMAN, M., KULIS, M., TAMBORERO, D., GOUIN, A., BLANC, J., GUT, M., GUT, I., PUENTE, X. S., PISANO,

- D. G., MARTIN-SUBERO, J. I., LOPEZ-BIGAS, N., LOPEZ-GUILLERMO, A., VALENCIA, A., LOPEZ-OTIN, C., CAMPO, E. & GUIGO, R. 2014. Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res*, 24, 212-26.
- FLYNN, R. L., COX, K. E., JEITANY, M., WAKIMOTO, H., BRYLL, A. R., GANEM, N. J., BERSANI, F., PINEDA, J. R., SUVA, M. L., BENES, C. H., HABER, D. A., BOUSSIN, F. D. & ZOU, L. 2015. Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science*, 347, 273-7.
- FORMENT, J. V., KAIDI, A. & JACKSON, S. P. 2012. Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nat Rev Cancer*, 12, 663-70.
- FRAENKEL, S., MOSTOSLAVSKY, R., NOVOBRANTSEVA, T. I., PELANDA, R., CHAUDHURI, J., ESPOSITO, G., JUNG, S., ALT, F. W., RAJEWSKY, K., CEDAR, H. & BERGMAN, Y. 2007. Allelic 'choice' governs somatic hypermutation *in vivo* at the immunoglobulin kappa-chain locus. *Nat Immunol*, 8, 715-22.
- FREEMAN, C. L. & GRIBBEN, J. G. 2016. Immunotherapy in Chronic Lymphocytic Leukaemia (CLL). *Curr Hematol Malig Rep*, 11, 29-36.
- FREITAS-JUNIOR, L. H., BOTTIUS, E., PIRRI, L. A., DEITSCH, K. W., SCHEIDIG, C., GUINET, F., NEHRBASS, U., WELLEMS, T. E. & SCHERF, A. 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature*, 407, 1018-22.
- FU, D., CALVO, J. A. & SAMSON, L. D. 2012. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer*, 12, 104-20.
- FUNGTAMMASAN, A., WALSH, E., CHIAROMONTE, F., ECKERT, K. A. & MAKOVA, K. D. 2012. A genome-wide analysis of common fragile sites: what features determine chromosomal instability in the human genome? *Genome Res*, 22, 993-1005.
- FUTREAL, P. A., COIN, L., MARSHALL, M., DOWN, T., HUBBARD, T., WOOSTER, R., RAHMAN, N. & STRATTON, M. R. 2004. A census of human cancer genes. *Nat Rev Cancer*, 4, 177-83.
- GARDINER, A. C., CORCORAN, M. M. & OSCIER, D. G. 1997. Cytogenetic, fluorescence *in situ* hybridisation, and clinical evaluation of translocations with concomitant deletion at 13q14 in chronic lymphocytic leukaemia. *Genes Chromosomes Cancer*, 20, 73-81.
- GARRAWAY, L. A. & LANDER, E. S. 2013. Lessons from the cancer genome. *Cell*, 153, 17-37.
- GENOMES PROJECT, C., AUTON, A., BROOKS, L. D., DURBIN, R. M., GARRISON, E. P., KANG, H. M., KORBEL, J. O., MARCHINI, J. L., MCCARTHY, S., MCVEAN, G. A. & ABECASIS, G. R. 2015. A global reference for human genetic variation. *Nature*, 526, 68-74.
- GHAMLOUCH, H., NGUYEN-KHAC, F. & BERNARD, O. A. 2017. Chronic lymphocytic leukaemia genomics and the precision medicine era. *Br J Haematol*.
- GHELLI LUSERNA DI RORA, A., IACOBUCCI, I. & MARTINELLI, G. 2017. The cell cycle checkpoint inhibitors in the treatment of leukemias. *J Hematol Oncol*, 10, 77.
- GHEZRAOUI, H., PIGANEAU, M., RENOUF, B., RENAUD, J. B., SALLMYR, A., RUIS, B., OH, S., TOMKINSON, A. E., HENDRICKSON, E. A., GIOVANNANGELI, C., JASIN, M. & BRUNET, E. 2014. Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol Cell*, 55, 829-42.
- GIBBS, K. D., JR., JAGER, A., CRESPO, O., GOLTSEV, Y., TREJO, A., RICHARD, C. E. & NOLAN, G. P. 2012. Decoupling of tumor-initiating activity from stable immunophenotype in HoxA9-Meis1-driven AML. *Cell Stem Cell*, 10, 210-7.
- GLADYCH, M., WOJTYLA, A. & RUBIS, B. 2011. Human telomerase expression regulation. *Biochem Cell Biol*, 89, 359-76.
- GOODARZI, A. A., YU, Y., RIBALLO, E., DOUGLAS, P., WALKER, S. A., YE, R., HARER, C., MARCHETTI, C., MORRICE, N., JEGGO, P. A. & LEES-MILLER, S. P. 2006. DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO J*, 25, 3880-9.

- GRAAKJAER, J., BISCHOFF, C., KORSHOLM, L., HOLSTEBROE, S., VACH, W., BOHR, V. A., CHRISTENSEN, K. & KOLVRAA, S. 2003. The pattern of chromosome-specific variations in telomere length in humans is determined by inherited, telomere-near factors and is maintained throughout life. *Mech Ageing Dev*, 124, 629-40.
- GRABOWSKI, P., HULTDIN, M., KARLSSON, K., TOBIN, G., ALESKOG, A., THUNBERG, U., LAURELL, A., SUNDSTROM, C., ROSENQUIST, R. & ROOS, G. 2005. Telomere length as a prognostic parameter in chronic lymphocytic leukemia with special reference to VH gene mutation status. *Blood*, 105, 4807-12.
- GREAVES, M. & MALEY, C. C. 2012. Clonal evolution in cancer. *Nature*, 481, 306-13.
- GREIDER, C. W. 1999. Telomeres do D-loop-T-loop. *Cell*, 97, 419-22.
- GREIDER, C. W. 2016. Regulating telomere length from the inside out: the replication fork model. *Genes Dev*, 30, 1483-91.
- GREIDER, C. W. & BLACKBURN, E. H. 1985. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell*, 43, 405-13.
- GREIDER, C. W. & BLACKBURN, E. H. 1989. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature*, 337, 331-7.
- GRIBBEN, J. G. 2009. Stem cell transplantation in chronic lymphocytic leukemia. *Biol Blood Marrow Transplant*, 15, 53-8.
- GRIFFITH, J. D., COMEAU, L., ROSENFELD, S., STANSEL, R. M., BIANCHI, A., MOSS, H. & DE LANGE, T. 1999. Mammalian telomeres end in a large duplex loop. *Cell*, 97, 503-14.
- GRYGALOWICZ, B., WORONIECKA, R., RYGIER, J., BORKOWSKA, K., RZEPECKA, I., LUKASIK, M., BUDZILOWSKA, A., RYMKIEWICZ, G., BLACHNIO, K., NOWAKOWSKA, B., BARTNIK, M., GOS, M. & PIENKOWSKA-GRELA, B. 2016. Monoallelic and biallelic deletions of 13q14 in a group of CLL/SLL patients investigated by CGH Haematological Cancer and SNP array (8x60K). *Mol Cytogenet*, 9, 1.
- GU, J., LI, S., ZHANG, X., WANG, L. C., NIEWOLIK, D., SCHWARZ, K., LEGERSKI, R. J., ZANDI, E. & LIEBER, M. R. 2010. DNA-PKcs regulates a single-stranded DNA endonuclease activity of Artemis. *DNA Repair (Amst)*, 9, 429-37.
- GU, P., WANG, Y., BISHT, K. K., WU, L., KUKOVA, L., SMITH, E. M., XIAO, Y., BAILEY, S. M., LEI, M., NANDAKUMAR, J. & CHANG, S. 2017. Pot1 OB-fold mutations unleash telomere instability to initiate tumorigenesis. *Oncogene*, 36, 1939-1951.
- GUARINI, A., GAIDANO, G., MAURO, F. R., CAPELLO, D., MANCINI, F., DE PROPRIIS, M. S., MANCINI, M., ORSINI, E., GENTILE, M., BRECCIA, M., CUNEO, A., CASTOLDI, G. & FOA, R. 2003. Chronic lymphocytic leukemia patients with highly stable and indolent disease show distinctive phenotypic and genotypic features. *Blood*, 102, 1035-41.
- GUEx, N. & PEITSCH, M. C. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 18, 2714-23.
- GUIEZE, R., PAGES, M., VERONESE, L., COMBES, P., LEMAL, R., GAY-BELLILE, M., CHAUVET, M., CALLANAN, M., KWIATKOWSKI, F., PEREIRA, B., VAGO, P., BAY, J. O., TOURNILHAC, O. & TCHIRKOV, A. 2016. Telomere status in chronic lymphocytic leukemia with TP53 disruption. *Oncotarget*, 7, 56976-56985.
- GUIEZE, R. & WU, C. J. 2015. Genomic and epigenomic heterogeneity in chronic lymphocytic leukemia. *Blood*, 126, 445-53.
- HAMBLIN, T. J., DAVIS, Z., GARDINER, A., OSCIER, D. G. & STEVENSON, F. K. 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*, 94, 1848-54.
- HARLEY, C. B., FUTCHER, A. B. & GREIDER, C. W. 1990. Telomeres shorten during ageing of human fibroblasts. *Nature*, 345, 458-60.

- 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, 866-8.
- HAYASHI, M. T., CESARE, A. J., RIVERA, T. & KARLSEDER, J. 2015. Cell death during crisis is mediated by mitotic telomere deprotection. *Nature*, 522, 492-6.
- HAYFLICK, L. 1965. The Limited *in vitro* Lifetime of Human Diploid Cell Strains. *Exp Cell Res*, 37, 614-36.
- HAYFLICK, L. & MOORHEAD, P. S. 1961. The serial cultivation of human diploid cell strains. *Exp Cell Res*, 25, 585-621.
- HEIDENREICH, B. & KUMAR, R. 2017. TERT promoter mutations in telomere biology. *Mutat Res*, 771, 15-31.
- HEMANN, M. T. & GREIDER, C. W. 2000. Wild-derived inbred mouse strains have short telomeres. *Nucleic Acids Res*, 28, 4474-8.
- HENSON, J. D., HANNAY, J. A., MCCARTHY, S. W., ROYDS, J. A., YEAGER, T. R., ROBINSON, R. A., WHARTON, S. B., JELLINEK, D. A., ARBUCKLE, S. M., YOO, J., ROBINSON, B. G., LEAROYD, D. L., STALLEY, P. D., BONAR, S. F., YU, D., POLLOCK, R. E. & REDDEL, R. R. 2005. A robust assay for alternative lengthening of telomeres in tumors shows the significance of alternative lengthening of telomeres in sarcomas and astrocytomas. *Clin Cancer Res*, 11, 217-25.
- HENSON, J. D., NEUMANN, A. A., YEAGER, T. R. & REDDEL, R. R. 2002. Alternative lengthening of telomeres in mammalian cells. *Oncogene*, 21, 598-610.
- HEUSER, M., YUN, H., BERG, T., YUNG, E., ARGIROPOULOS, B., KUCHENBAUER, F., PARK, G., HAMWI, I., PALMQVIST, L., LAI, C. K., LEUNG, M., LIN, G., CHATURVEDI, A., THAKUR, B. K., IWASAKI, M., BILENKY, M., THIESSEN, N., ROBERTSON, G., HIRST, M., KENT, D., WILSON, N. K., GOTTGENS, B., EAVES, C., CLEARY, M. L., MARRA, M., GANSER, A. & HUMPHRIES, R. K. 2011. Cell of origin in AML: susceptibility to MN1-induced transformation is regulated by the MEIS1/AbdB-like HOX protein complex. *Cancer Cell*, 20, 39-52.
- HIGGINS, G. S., HARRIS, A. L., PREVO, R., HELLEDAY, T., MCKENNA, W. G. & BUFFA, F. M. 2010. Overexpression of POLQ confers a poor prognosis in early breast cancer patients. *Oncotarget*, 1, 175-84.
- HILMI, K., JANGAL, M., MARQUES, M., ZHAO, T., SAAD, A., ZHANG, C., LUO, V. M., SYME, A., REJON, C., YU, Z., KRUM, A., FABIAN, M. R., RICHARD, S., ALAOUJ-JAMALI, M., ORTHWEIN, A., MCCAFFREY, L. & WITCHER, M. 2017. CTCF facilitates DNA double-strand break repair by enhancing homologous recombination repair. *Sci Adv*, 3, e1601898.
- HLADILKOVA, E., BAROY, T., FANNEMEL, M., VALLOVA, V., MISCEO, D., BRYN, V., SLAMOVA, I., PRASILOVA, S., KUGLIK, P. & FRENGEN, E. 2015. A recurrent deletion on chromosome 2q13 is associated with developmental delay and mild facial dysmorphisms. *Mol Cytogenet*, 8, 57.
- HO, A., WILSON, F. R., PERAGINE, S. L., JEYANTHAN, K., MITCHELL, T. R. & ZHU, X. D. 2016. TRF1 phosphorylation on T271 modulates telomerase-dependent telomere length maintenance as well as the formation of ALT-associated PML bodies. *Sci Rep*, 6, 36913.
- HOANG, D., SUE, G. R., XU, F., LI, P. & NARAYAN, D. 2013. Absence of aneuploidy and gastrointestinal tumours in a man with a chromosomal 2q13 deletion and BUB1 monoallelic deficiency. *BMJ Case Rep*, 2013.
- HOCK, B. D., FERNYHOUGH, L. J., GOUGH, S. M., STEINKASSERER, A., COX, A. G. & MCKENZIE, J. L. 2009. Release and clinical significance of soluble CD83 in chronic lymphocytic leukemia. *Leuk Res*, 33, 1089-95.

- HODES, R. J. 1999. Telomere length, aging, and somatic cell turnover. *J Exp Med*, 190, 153-6.
- HOFMAN, I. J. F., VAN DUIN, M., DE BRUYNE, E., FANCELLO, L., MULLIGAN, G., GEERDENS, E., GARELLI, E., MANCINI, C., LEMMENS, H., DELFORGE, M., VANDENBERGHE, P., WLODARSKA, I., ASPESI, A., MICHAUX, L., VANDERKERKEN, K., SONNEVELD, P. & DE KEERSMAECKER, K. 2017. RPL5 on 1p22.1 is recurrently deleted in multiple myeloma and its expression is linked to bortezomib response. *Leukemia*, 31, 1706-1714.
- HOLLAND, A. J. & CLEVELAND, D. W. 2012. Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. *Nat Med*, 18, 1630-8.
- HORN, S., FIGL, A., RACHAKONDA, P. S., FISCHER, C., SUCKER, A., GAST, A., KADEL, S., MOLL, I., NAGORE, E., HEMMINKI, K., SCHADENDORF, D. & KUMAR, R. 2013. TERT promoter mutations in familial and sporadic melanoma. *Science*, 339, 959-61.
- HOWARD, D. R., MUNIR, T., MCPARLAND, L., RAWSTRON, A. C., MILLIGAN, D., SCHUH, A., HOCKADAY, A., ALLSUP, D. J., MARSHALL, S., DUNCOMBE, A. S., O'DWYER, J. L., SMITH, A. F., LONGO, R., VARGHESE, A. & HILLMEN, P. 2017. Results of the randomized phase IIB ARCTIC trial of low-dose rituximab in previously untreated CLL. *Leukemia*.
- HOWLADER N, N. A., KRAPCHO M, MILLER D, BISHOP K, KOSARY CL, YU M, RUHL J, TATALOVICH Z, MARIOTTO A, LEWIS DR, CHEN HS, FEUER EJ, CRONIN KA (EDS). 2017. SEER Cancer Statistics Review (CSR) 1975-2014. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2014/, based on November 2016 SEER data submission, posted to the SEER web site, April 2017.
- HRUBA, M., DVORAK, P., WEBEROVA, L. & SUBRT, I. 2012. Independent coexistence of clones with 13q14 deletion at reciprocal translocation breakpoint and 13q14 interstitial deletion in chronic lymphocytic leukemia. *Leuk Lymphoma*, 53, 2054-62.
- HU, B. T., LEE, S. C., MARIN, E., RYAN, D. H. & INSEL, R. A. 1997. Telomerase is up-regulated in human germinal center B cells *in vivo* and can be re-expressed in memory B cells activated *in vitro*. *J Immunol*, 159, 1068-71.
- HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37, 1-13.
- HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4, 44-57.
- HUANG, F. W., HODIS, E., XU, M. J., KRYUKOV, G. V., CHIN, L. & GARRAWAY, L. A. 2013. Highly recurrent TERT promoter mutations in human melanoma. *Science*, 339, 957-9.
- HUANG, K. K., JANG, K. W., KIM, S., KIM, H. S., KIM, S. M., KWON, H. J., KIM, H. R., YUN, H. J., AHN, M. J., PARK, K. U., RAMNARAYANAN, K., MCPHERSON, J. R., ZHANG, S., RHEE, J. K., VETTORE, A. L., DAS, K., ISHIMOTO, T., KIM, J. H., KOH, Y. W., KIM, S. H., CHOI, E. C., TEH, B. T., ROZEN, S. G., KIM, T. M., TAN, P. & CHO, B. C. 2016. Exome sequencing reveals recurrent REV3L mutations in cisplatin-resistant squamous cell carcinoma of head and neck. *Sci Rep*, 6, 19552.
- HUFFMAN, K. E., LEVENE, S. D., TESMER, V. M., SHAY, J. W. & WRIGHT, W. E. 2000. Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang. *J Biol Chem*, 275, 19719-22.
- HULTDIN, M., ROSENQUIST, R., THUNBERG, U., TOBIN, G., NORRBACK, K. F., JOHNSON, A., SUNDSTROM, C. & ROOS, G. 2003. Association between telomere length and V(H)

- gene mutation status in chronic lymphocytic leukaemia: clinical and biological implications. *Br J Cancer*, 88, 593-8.
- HUS, I., PODHORECKA, M., BOJARSKA-JUNAK, A., ROLINSKI, J., SCHMITT, M., SIEKLUCKA, M., WASIK-SZCZEPANEK, E. & DMOSZYNSKA, A. 2006. The clinical significance of ZAP-70 and CD38 expression in B-cell chronic lymphocytic leukaemia. *Ann Oncol*, 17, 683-90.
- HWANG, H., BUNCHER, N., OPRESKO, P. L. & MYONG, S. 2012. POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure*, 20, 1872-80.
- HYATT, S., JONES, R. E., HEPPEL, N. H., GRIMSTEAD, J. W., FEGAN, C., JACKSON, G. H., HILLS, R., ALLAN, J. M., PRATT, G., PEPPER, C. & BAIRD, D. M. 2017. Telomere length is a critical determinant for survival in multiple myeloma. *Br J Haematol*, 178, 94-98.
- IJDO, J. W., BALDINI, A., WARD, D. C., REEDERS, S. T. & WELLS, R. A. 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci U S A*, 88, 9051-5.
- ILIAKIS, G., MURMANN, T. & SONI, A. 2015. Alternative end-joining repair pathways are the ultimate backup for abrogated classical non-homologous end-joining and homologous recombination repair: Implications for the formation of chromosome translocations. *Mutat Res Genet Toxicol Environ Mutagen*, 793, 166-75.
- ISHDORJ, G., KOST, S. E. F., BEIGGI, S., ZANG, Y., GIBSON, S. B. & JOHNSTON, J. B. 2017. A novel spliced variant of the TIN2 shelterin is present in chronic lymphocytic leukemia. *Leuk Res*, 59, 66-74.
- ISONO, M., NIIMI, A., OIKE, T., HAGIWARA, Y., SATO, H., SEKINE, R., YOSHIDA, Y., ISOBE, S. Y., OBUSE, C., NISHI, R., PETRICCI, E., NAKADA, S., NAKANO, T. & SHIBATA, A. 2017. BRCA1 Directs the Repair Pathway to Homologous Recombination by Promoting 53BP1 Dephosphorylation. *Cell Rep*, 18, 520-532.
- JACKSON, S. P. & BARTEK, J. 2009. The DNA-damage response in human biology and disease. *Nature*, 461, 1071-8.
- JACOB, B., OSATO, M., YAMASHITA, N., WANG, C. Q., TANIUCHI, I., LITTMAN, D. R., ASOU, N. & ITO, Y. 2010. Stem cell exhaustion due to Runx1 deficiency is prevented by Evi5 activation in leukemogenesis. *Blood*, 115, 1610-20.
- JACOB, N. K., STOUT, A. R. & PRICE, C. M. 2004. Modulation of telomere length dynamics by the subtelomeric region of tetrahymena telomeres. *Mol Biol Cell*, 15, 3719-28.
- JACOBY, M. A., DUNCAVAGE, E. J. & WALTER, M. J. 2015. Implications of Tumor Clonal Heterogeneity in the Era of Next-Generation Sequencing. *Trends Cancer*, 1, 231-241.
- JAIN, P. & O'BRIEN, S. 2012. Richter's transformation in chronic lymphocytic leukemia. *Oncology (Williston Park)*, 26, 1146-52.
- JASEK, M., GONDEK, L. P., BEJANYAN, N., TIU, R., HUH, J., THEIL, K. S., O'KEEFE, C., MCDEVITT, M. A. & MACIEJEWSKI, J. P. 2010. TP53 mutations in myeloid malignancies are either homozygous or hemizygous due to copy number-neutral loss of heterozygosity or deletion of 17p. *Leukemia*, 24, 216-9.
- JITSCHIN, R., HOFMANN, A. D., BRUNS, H., GIESSL, A., BRICKS, J., BERGER, J., SAUL, D., ECKART, M. J., MACKENSEN, A. & MOUGIAKAKOS, D. 2014. Mitochondrial metabolism contributes to oxidative stress and reveals therapeutic targets in chronic lymphocytic leukemia. *Blood*, 123, 2663-72.
- 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 telomere-driven crisis is DNA ligase III dependent. *Cell Rep*, 8, 1063-76.
- JOZA, N., SUSIN, S. A., DAUGAS, E., STANFORD, W. L., CHO, S. K., LI, C. Y., SASAKI, T., ELIA, A. J., CHENG, H. Y., RAVAGNAN, L., FERRI, K. F., ZAMZAMI, N., WAKEHAM, A., HAKEM, R., YOSHIDA, H., KONG, Y. Y., MAK, T. W., ZUNIGA-PFLUCKER, J. C., KROEMER, G. &

- PENNINGER, J. M. 2001. Essential role of the mitochondrial apoptosis-inducing factor in programmed cell death. *Nature*, 410, 549-54.
- JU, Y. S., MARTINCORENA, I., GERSTUNG, M., PETLJAK, M., ALEXANDROV, L. B., RAHBARI, R., WEDGE, D. C., DAVIES, H. R., RAMAKRISHNA, M., FULLAM, A., MARTIN, S., ALDER, C., PATEL, N., GAMBLE, S., O'MEARA, S., GIRI, D. D., SAUER, T., PINDER, S. E., PURDIE, C. A., BORG, A., STUNNENBERG, H., VAN DE VIJVER, M., TAN, B. K., CALDAS, C., TUTT, A., UENO, N. T., VAN 'T VEER, L. J., MARTENS, J. W., SOTIRIOU, C., KNAPPSKOG, S., SPAN, P. N., LAKHANI, S. R., EYFJORD, J. E., BORRESEN-DALE, A. L., RICHARDSON, A., THOMPSON, A. M., VIARI, A., HURLES, M. E., NIK-ZAINAL, S., CAMPBELL, P. J. & STRATTON, M. R. 2017. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature*, 543, 714-718.
- JU, Y. S., TUBIO, J. M., MIFSUD, W., FU, B., DAVIES, H. R., RAMAKRISHNA, M., LI, Y., YATES, L., GUNDEM, G., TARPEY, P. S., BEHJATI, S., PAPAEMMANUIL, E., MARTIN, S., FULLAM, A., GERSTUNG, M., GROUP, I. P. C. W., GROUP, I. B. C. W., GROUP, I. B. C. W., NANGALIA, J., GREEN, A. R., CALDAS, C., BORG, A., TUTT, A., LEE, M. T., VAN'T VEER, L. J., TAN, B. K., APARICIO, S., SPAN, P. N., MARTENS, J. W., KNAPPSKOG, S., VINCENT-SALOMON, A., BORRESEN-DALE, A. L., EYFJORD, J. E., MYKLEBOST, O., FLANAGAN, A. M., FOSTER, C., NEAL, D. E., COOPER, C., EELES, R., BOVA, S. G., LAKHANI, S. R., DESMEDT, C., THOMAS, G., RICHARDSON, A. L., PURDIE, C. A., THOMPSON, A. M., MCDERMOTT, U., YANG, F., NIK-ZAINAL, S., CAMPBELL, P. J. & STRATTON, M. R. 2015. Frequent somatic transfer of mitochondrial DNA into the nuclear genome of human cancer cells. *Genome Res*, 25, 814-24.
- KAKAROUGKAS, A., ISMAIL, A., CHAMBERS, A. L., RIBALLO, E., HERBERT, A. D., KUNZEL, J., LOBRICH, M., JEGGO, P. A. & DOWNS, J. A. 2014. Requirement for PBAF in transcriptional repression and repair at DNA breaks in actively transcribed regions of chromatin. *Mol Cell*, 55, 723-32.
- KAMIGUTI, A. S., SERRANDER, L., LIN, K., HARRIS, R. J., CAWLEY, J. C., ALLSUP, D. J., SLUPSKY, J. R., KRAUSE, K. H. & ZUZEL, M. 2005. Expression and Activity of NOX5 in the Circulating Malignant B Cells of Hairy Cell Leukemia. *The Journal of Immunology*, 175, 8424-8430.
- KANDASWAMY, R., SAVA, G. P., SPEEDY, H. E., BEA, S., MARTIN-SUBERO, J. I., STUDD, J. B., MIGLIORINI, G., LAW, P. J., PUENTE, X. S., MARTIN-GARCIA, D., SALAVERRIA, I., GUTIERREZ-ABRIL, J., LOPEZ-OTIN, C., CATOVSKY, D., ALLAN, J. M., CAMPO, E. & HOULSTON, R. S. 2016. Genetic Predisposition to Chronic Lymphocytic Leukemia Is Mediated by a BMF Super-Enhancer Polymorphism. *Cell Rep*, 16, 2061-7.
- KARLSEDER, J., HOKE, K., MIRZOEVA, O. K., BAKKENIST, C., KASTAN, M. B., PETRINI, J. H. & DE LANGE, T. 2004. The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response. *PLoS Biol*, 2, E240.
- KASAR, S., KIM, J., IMPROGO, R., TIAO, G., POLAK, P., HARADHVALA, N., LAWRENCE, M. S., KIEZUN, A., FERNANDES, S. M., BAHL, S., SOUGNEZ, C., GABRIEL, S., LANDER, E. S., KIM, H. T., GETZ, G. & BROWN, J. R. 2015. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun*, 6, 8866.
- KENT, T., CHANDRAMOULY, G., MCDEVITT, S. M., OZDEMIR, A. Y. & POMERANTZ, R. T. 2015. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat Struct Mol Biol*, 22, 230-7.
- KERN, W., BACHER, U., HAFERLACH, C., ALPERMANN, T., DICKER, F., SCHNITTGER, S. & HAFERLACH, T. 2012. Frequency and prognostic impact of the aberrant CD8 expression in 5,523 patients with chronic lymphocytic leukemia. *Cytometry B Clin Cytom*, 82, 145-50.

- KIM, C., BASSIG, B. A., SEOW, W. J., HU, W., PURDUE, M. P., HUANG, W. Y., LIU, C. S., CHENG, W. L., MANNISTO, S., VERMEULEN, R., WEINSTEIN, S. J., LIM, U., HOSGOOD, H. D., BONNER, M. R., CAPORASO, N. E., ALBANES, D., LAN, Q. & ROTHMAN, N. 2015. Mitochondrial DNA copy number and chronic lymphocytic leukemia/small lymphocytic lymphoma risk in two prospective studies. *Cancer Epidemiol Biomarkers Prev*, 24, 148-53.
- 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*, 266, 2011-5.
- KIM, W., LUDLOW, A. T., MIN, J., ROBIN, J. D., STADLER, G., MENDER, I., LAI, T. P., ZHANG, N., WRIGHT, W. E. & SHAY, J. W. 2016. Regulation of the Human Telomerase Gene TERT by Telomere Position Effect-Over Long Distances (TPE-OLD): Implications for Aging and Cancer. *PLoS Biol*, 14, e2000016.
- 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. *Nat Protoc*, 5, 1596-607.
- KIPLING, D. & COOKE, H. J. 1990. Hypervariable ultra-long telomeres in mice. *Nature*, 347, 400-2.
- KIPPS, T. J., STEVENSON, F. K., WU, C. J., CROCE, C. M., PACKHAM, G., WIERDA, W. G., O'BRIEN, S., GRIBBEN, J. & RAI, K. 2017. Chronic lymphocytic leukaemia. *Nat Rev Dis Primers*, 3, 16096.
- KLAMBAUER, G., SCHWARZBAUER, K., MAYR, A., CLEVERT, D. A., MITTERECKER, A., BODENHOFER, U. & HOCHREITER, S. 2012. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res*, 40, e69.
- KLEIN, U., LIA, M., CRESPO, M., SIEGEL, R., SHEN, Q., MO, T., AMBESI-IMPIOMBATO, A., CALIFANO, A., MIGLIAZZA, A., BHAGAT, G. & DALLA-FAVERA, R. 2010. The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. *Cancer Cell*, 17, 28-40.
- KNITTEL, G., REHKAMPER, T., KOROVKINA, D., LIEDGENS, P., FRITZ, C., TORGOVNICK, A., ALBALDAWI, Y., AL-MAARRI, M., CUN, Y., FEDORCHENKO, O., RIABINSKA, A., BELEGGIA, F., NGUYEN, P. H., WUNDERLICH, F. T., ORTMANN, M., MONTESINOS-RONGEN, M., TAUSCH, E., STILGENBAUER, S., L, P. F., HERLING, M., HERLING, C., BAHLO, J., HALLEK, M., PEIFER, M., BUETTNER, R., PERSIGHEHL, T. & REINHARDT, H. C. 2017. Two mouse models reveal an actionable PARP1 dependence in aggressive chronic lymphocytic leukemia. *Nat Commun*, 8, 153.
- KOCZULA, K. M., LUDWIG, C., HAYDEN, R., CRONIN, L., PRATT, G., PARRY, H., TENNANT, D., DRAYSON, M., BUNCE, C. M., KHANIM, F. L. & GUNTHER, U. L. 2016. Metabolic plasticity in CLL: adaptation to the hypoxic niche. *Leukemia*, 30, 65-73.
- KOO, D. H., SINGH, B., JIANG, J., FRIEBE, B., GILL, B. S., CHASTAIN, P. D., MANNE, U., TIWARI, H. K. & SINGH, K. K. 2017. Single molecule mtDNA fiber FISH for analyzing numtogenesis. *Anal Biochem*.
- KREJCI, K. & KOCH, J. 1998. Improved detection and comparative sizing of human chromosomal telomeres *in situ*. *Chromosoma*, 107, 198-203.
- KRETSCHMER, B., LUTHJE, K., SCHNEIDER, S., FLEISCHER, B. & BRELOER, M. 2009. Engagement of CD83 on B cells modulates B cell function *in vivo*. *J Immunol*, 182, 2827-34.
- KRETZOVALI, A., AGALITI, T., SPILIANAKIS, C., TZORTZAKAKI, E., MERIKA, M. & PAPAMATHEAKIS, J. 1998. Involvement of CREB binding protein in expression of

- major histocompatibility complex class II genes via interaction with the class II transactivator. *Mol Cell Biol*, 18, 6777-83.
- KRIANGKUM, J., MOTZ, S. N., MACK, T., BEIGGI, S., BAIGORRI, E., KUPPUSAMY, H., BELCH, A. R., JOHNSTON, J. B. & PILARSKI, L. M. 2015. Single-Cell Analysis and Next-Generation Immuno-Sequencing Show That Multiple Clones Persist in Patients with Chronic Lymphocytic Leukemia. *PLoS One*, 10, e0137232.
- KROBER, A., SEILER, T., BENNER, A., BULLINGER, L., BRUCKLE, E., LICHTER, P., DOHNER, H. & STILGENBAUER, S. 2002. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood*, 100, 1410-6.
- KRZYWINSKI, M., SCHEIN, J., BIROL, I., CONNORS, J., GASCOYNE, R., HORSMAN, D., JONES, S. J. & MARRA, M. A. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res*, 19, 1639-45.
- KULIS, M., HEATH, S., BIBIKOVA, M., QUEIROS, A. C., NAVARRO, A., CLOT, G., MARTINEZ-TRILLOS, A., CASTELLANO, G., BRUN-HEATH, I., PINYOL, M., BARBERAN-SOLER, S., PAPASAIKAS, P., JARES, P., BEA, S., RICO, D., ECKER, S., RUBIO, M., ROYO, R., HO, V., KLOTZLE, B., HERNANDEZ, L., CONDE, L., LOPEZ-GUERRA, M., COLOMER, D., VILLAMOR, N., AYMERICH, M., ROZMAN, M., BAYES, M., GUT, M., GELPI, J. L., OROZCO, M., FAN, J. B., QUESADA, V., PUENTE, X. S., PISANO, D. G., VALENCIA, A., LOPEZ-GUILLERMO, A., GUT, I., LOPEZ-OTIN, C., CAMPO, E. & MARTIN-SUBERO, J. I. 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet*, 44, 1236-42.
- KUMAR, P., HENIKOFF, S. & NG, P. C. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 4, 1073-81.
- KYO, S., TAKAKURA, M., FUJIWARA, T. & INOUE, M. 2008. Understanding and exploiting hTERT promoter regulation for diagnosis and treatment of human cancers. *Cancer Sci*, 99, 1528-38.
- LANDAU, D. A., CARTER, S. L., GETZ, G. & WU, C. J. 2014. Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia*, 28, 34-43.
- LANDAU, D. A., CARTER, S. L., STOJANOV, P., MCKENNA, A., STEVENSON, K., LAWRENCE, M. S., SOUGNEZ, C., STEWART, C., SIVACHENKO, A., WANG, L., WAN, Y., ZHANG, W., SHUKLA, S. A., VARTANOV, A., FERNANDES, S. M., SAKSENA, G., CIBULSKIS, K., TESAR, B., GABRIEL, S., HACOEN, N., MEYERSON, M., LANDER, E. S., NEUBERG, D., BROWN, J. R., GETZ, G. & WU, C. J. 2013. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*, 152, 714-26.
- LANDAU, D. A., TAUSCH, E., TAYLOR-WEINER, A. N., STEWART, C., REITER, J. G., BAHLO, J., KLUTH, S., BOZIC, I., LAWRENCE, M., BOTTCHER, S., CARTER, S. L., CIBULSKIS, K., MERTENS, D., SOUGNEZ, C. L., ROSENBERG, M., HESS, J. M., EDELMANN, J., KLESS, S., KNEBA, M., RITGEN, M., FINK, A., FISCHER, K., GABRIEL, S., LANDER, E. S., NOWAK, M. A., DOHNER, H., HALLEK, M., NEUBERG, D., GETZ, G., STILGENBAUER, S. & WU, C. J. 2015. Mutations driving CLL and their evolution in progression and relapse. *Nature*, 526, 525-30.
- LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGETHOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S.,

- HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
- LANGE, S. S., BEDFORD, E., REH, S., WITTSCHIEBEN, J. P., CARBAJAL, S., KUSEWITT, D. F., DIGIOVANNI, J. & WOOD, R. D. 2013. Dual role for mammalian DNA polymerase zeta in maintaining genome stability and proliferative responses. *Proc Natl Acad Sci U S A*, 110, E687-96.
- LANGE, S. S., TOMIDA, J., BOULWARE, K. S., BHETAWAL, S. & WOOD, R. D. 2016. The Polymerase Activity of Mammalian DNA Pol zeta Is Specifically Required for Cell and Embryonic Viability. *PLoS Genet*, 12, e1005759.
- 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. *Hum Mol Genet*, 5, 685-91.
- LARSON, D. E., HARRIS, C. C., CHEN, K., KOBOLDT, D. C., ABBOTT, T. E., DOOLING, D. J., LEY, T. J., MARDIS, E. R., WILSON, R. K. & DING, L. 2012. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*, 28, 311-7.
- LAW, P. J., BERNDT, S. I., SPEEDY, H. E., CAMP, N. J., SAVA, G. P., SKIBOLA, C. F., HOLROYD, A., JOSEPH, V., SUNTER, N. J., NIETERS, A., BEA, S., MONNEREAU, A., MARTIN-GARCIA, D., GOLDIN, L. R., CLOT, G., TERAS, L. R., QUINTELA, I., BIRMAN, B. M., JAYNE, S., COZEN, W., MAJID, A., SMEDBY, K. E., LAN, Q., DEARDEN, C., BROOKS-WILSON, A. R., HALL, A. G., PURDUE, M. P., MAINOU-FOWLER, T., VAJDIC, C. M., JACKSON, G. H., COCCO, P., MARR, H., ZHANG, Y., ZHENG, T., GILES, G. G., LAWRENCE, C., CALL, T. G., LIEBOW, M., MELBYE, M., GLIMELIUS, B., MANSOURI, L., GLENN, M., CURTIN, K., DIVER, W. R., LINK, B. K., CONDE, L., BRACCI, P. M., HOLLY, E. A., JACKSON, R. D., TINKER, L. F., BENAVENTE, Y., BOFFETTA, P., BRENNAN, P., MAYNADIE, M., MCKAY, J., ALBANES, D., WEINSTEIN, S., WANG, Z., CAPORASO, N. E., MORTON, L. M., SEVERSON, R. K., RIBOLI, E., VINEIS, P., VERMEULEN, R. C., SOUTHEY, M. C., MILNE, R. L., CLAVEL, J., TOPKA, S., SPINELLI, J. J., KRAFT, P., ENNAS, M. G., SUMMERFIELD, G., FERRI, G. M., HARRIS, R. J., MILIGI, L., PETTITT, A. R., NORTH, K. E., ALLSUP, D. J., FRAUMENI, J. F., BAILEY, J. R., OFFIT, K., PRATT, G., HJALGRIM, H., PEPPER, C., CHANOCK, S. J., FEGAN, C., ROSENQUIST, R., DE SANJOSE, S., CARRACEDO, A., DYER, M. J., CATOVSKY, D., CAMPO, E., CERHAN, J. R., ALLAN, J. M., ROTHMAN, N., HOULSTON, R. & SLAGER, S. 2017. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. *Nat Commun*, 8, 14175.
- LAZZERINI-DENCHI, E. & SFEIR, A. 2016. Stop pulling my strings - what telomeres taught us about the DNA damage response. *Nat Rev Mol Cell Biol*, 17, 364-78.
- LEE, M., HILLS, M., CONOMOS, D., STUTZ, M. D., DAGG, R. A., LAU, L. M., REDDEL, R. R. & PICKETT, H. A. 2014a. Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. *Nucleic Acids Res*, 42, 1733-46.
- LEE, Y. S., GREGORY, M. T. & YANG, W. 2014b. Human Pol zeta purified with accessory subunits is active in translesion DNA synthesis and complements Pol eta in cisplatin bypass. *Proc Natl Acad Sci U S A*, 111, 2954-9.

- LEMEE, F., BERGOGLIO, V., FERNANDEZ-VIDAL, A., MACHADO-SILVA, A., PILLAIRE, M. J., BIETH, A., GENTIL, C., BAKER, L., MARTIN, A. L., LEDUC, C., LAM, E., MAGDELEINE, E., FILLERON, T., OUMOUHOU, N., KAINA, B., SEKI, M., GRIMAL, F., LACROIX-TRIKI, M., THOMPSON, A., ROCHE, H., BOURDON, J. C., WOOD, R. D., HOFFMANN, J. S. & CAZAUX, C. 2010. DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc Natl Acad Sci U S A*, 107, 13390-5.
- LETSOLO, B. T., ROWSON, J. & BAIRD, D. M. 2010. Fusion of short telomeres in human cells is characterized by extensive deletion and microhomology, and can result in complex rearrangements. *Nucleic Acids Res*, 38, 1841-52.
- LI, A. Y., LIN, H. H., KUO, C. Y., SHIH, H. M., WANG, C. C., YEN, Y. & ANN, D. K. 2011. High-mobility group A2 protein modulates hTERT transcription to promote tumorigenesis. *Mol Cell Biol*, 31, 2605-17.
- LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754-60.
- LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-9.
- LI, J. S., MIRALLES FUSTE, J., SIMAVORIAN, T., BARTOCCI, C., TSAI, J., KARLSEDER, J. & LAZZERINI DENCHI, E. 2017. TZAP: A telomere-associated protein involved in telomere length control. *Science*, 355, 638-641.
- LIDDIARD, K., RUIS, B., TAKASUGI, T., HARVEY, A., ASHELFORD, K. E., HENDRICKSON, E. A. & BAIRD, D. M. 2016. Sister chromatid telomere fusions, but not NHEJ-mediated inter-chromosomal telomere fusions, occur independently of DNA ligases 3 and 4. *Genome Res*, 26, 588-600.
- LIEBER, M. R. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*, 79, 181-211.
- LIEBER, M. R. 2016. Mechanisms of human lymphoid chromosomal translocations. *Nat Rev Cancer*, 16, 387-98.
- LIN, S.-Y. & ELLEDGE, S. J. 2003. Multiple Tumor Suppressor Pathways Negatively Regulate Telomerase. *Cell*, 113, 881-889.
- LIN, T. T., LETSOLO, B. T., JONES, R. E., ROWSON, J., PRATT, G., HEWAMANA, S., FEGAN, C., PEPPER, C. & BAIRD, D. M. 2010. Telomere dysfunction and fusion during the progression of chronic lymphocytic leukemia: evidence for a telomere crisis. *Blood*, 116, 1899-907.
- LIN, T. T., 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. *Br J Haematol*, 167, 214-23.
- LINARDOPOULOU, E. V., PARGHI, S. S., FRIEDMAN, C., OSBORN, G. E., PARKHURST, S. M. & TRASK, B. J. 2007. Human subtelomeric WASH genes encode a new subclass of the WASP family. *PLoS Genet*, 3, e237.
- LINARDOPOULOU, E. V., WILLIAMS, E. M., FAN, Y., FRIEDMAN, C., YOUNG, J. M. & TRASK, B. J. 2005. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature*, 437, 94-100.
- LIU, D., SAFARI, A., O'CONNOR, M. S., CHAN, D. W., LAEGELER, A., QIN, J. & SONGYANG, Z. 2004. PTPN22 interacts with POT1 and regulates its localization to telomeres. *Nat Cell Biol*, 6, 673-80.

- LIU, J., QIN, Y. Z., YANG, S., WANG, Y., CHANG, Y. J., ZHAO, T., JIANG, Q. & HUANG, X. J. 2017a. Meis1 is critical to the maintenance of human acute myeloid leukemia cells independent of MLL rearrangements. *Ann Hematol*, 96, 567-574.
- LIU, L., TRIMARCHI, J. R., SMITH, P. J. & KEEFE, D. L. 2002. Mitochondrial dysfunction leads to telomere attrition and genomic instability. *Aging Cell*, 1, 40-6.
- LIU, X., LI, F., HUANG, Q., ZHANG, Z., ZHOU, L., DENG, Y., ZHOU, M., FLEENOR, D. E., WANG, H., KASTAN, M. B. & LI, C. Y. 2017b. Self-inflicted DNA double-strand breaks sustain tumorigenicity and stemness of cancer cells. *Cell Res*, 27, 764-783.
- LIU, X. S. & MARDIS, E. R. 2017. Applications of Immunogenomics to Cancer. *Cell*, 168, 600-612.
- LIU, Y., WANG, R., ZHANG, L., LI, J., LOU, K. & SHI, B. 2017c. The lipid metabolism gene FTO influences breast cancer cell energy metabolism via the PI3K/AKT signaling pathway. *Oncol Lett*, 13, 4685-4690.
- LIVRAGHI, L. & GARBER, J. E. 2015. PARP inhibitors in the management of breast cancer: current data and future prospects. *BMC Med*, 13, 188.
- LO, A. W., SABATIER, L., FOULADI, B., POTTIER, G., RICOUL, M. & MURNANE, J. P. 2002. DNA amplification by breakage/fusion/bridge cycles initiated by spontaneous telomere loss in a human cancer cell line. *Neoplasia*, 4, 531-8.
- LOAYZA, D., PARSONS, H., DONIGIAN, J., HOKE, K. & DE LANGE, T. 2004. DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites. *J Biol Chem*, 279, 13241-8.
- LONDONO-VALLEJO, J. A., DER-SARKISSIAN, H., CAZES, L., BACCHETTI, S. & REDDEL, R. R. 2004. Alternative lengthening of telomeres is characterized by high rates of telomeric exchange. *Cancer Res*, 64, 2324-7.
- LONDONO-VALLEJO, J. A., DERSARKISSIAN, H., CAZES, L. & THOMAS, G. 2001. Differences in telomere length between homologous chromosomes in humans. *Nucleic Acids Res*, 29, 3164-71.
- LORD, C. J. & ASHWORTH, A. 2016. BRCAness revisited. *Nat Rev Cancer*, 16, 110-20.
- LOTTERSBERGER, F., BOTHMER, A., ROBBIANI, D. F., NUSSENZWEIG, M. C. & DE LANGE, T. 2013. Role of 53BP1 oligomerization in regulating double-strand break repair. *Proc Natl Acad Sci U S A*, 110, 2146-51.
- LOWDEN, M. R., FLIBOTTE, S., MOERMAN, D. G. & AHMED, S. 2011. DNA synthesis generates terminal duplications that seal end-to-end chromosome fusions. *Science*, 332, 468-71.
- LU, G., DUAN, J., SHU, S., WANG, X., GAO, L., GUO, J. & ZHANG, Y. 2016. Ligase I and ligase III mediate the DNA double-strand break ligation in alternative end-joining. *Proc Natl Acad Sci U S A*, 113, 1256-60.
- MA, C., MARTIN, S., TRASK, B. & HAMLIN, J. L. 1993. Sister chromatid fusion initiates amplification of the dihydrofolate reductase gene in Chinese hamster cells. *Genes Dev*, 7, 605-20.
- MA, Y., LU, H., TIPPIN, B., GOODMAN, M. F., SHIMAZAKI, N., KOIWAI, O., HSIEH, C. L., SCHWARZ, K. & LIEBER, M. R. 2004. A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol Cell*, 16, 701-13.
- MACHIELA, M. J., LAN, Q., SLAGER, S. L., VERMEULEN, R. C., TERAS, L. R., CAMP, N. J., CERHAN, J. R., SPINELLI, J. J., WANG, S. S., NIETERS, A., VIJAI, J., YEAGER, M., WANG, Z., GHESQUIERES, H., MCKAY, J., CONDE, L., DE BAKKER, P. I., COX, D. G., BURDETT, L., MONNEREAU, A., FLOWERS, C. R., DE ROOS, A. J., BROOKS-WILSON, A. R., GILES, G. G., MELBYE, M., GU, J., JACKSON, R. D., KANE, E., PURDUE, M. P., VAJDIC, C. M., ALBANES, D., KELLY, R. S., ZUCCA, M., BERTRAND, K. A., ZELENIUCH-JACQUOTTE, A., LAWRENCE, C., HUTCHINSON, A., ZHI, D., HABERMANN, T. M., LINK, B. K., NOVAK,

- A. J., DOGAN, A., ASMANN, Y. W., LIEBOW, M., THOMPSON, C. A., ANSELL, S. M., WITZIG, T. E., TILLY, H., HAIOUN, C., MOLINA, T. J., HJALGRIM, H., GLIMELIUS, B., ADAMI, H. O., ROOS, G., BRACCI, P. M., RIBY, J., SMITH, M. T., HOLLY, E. A., COZEN, W., HARTGE, P., MORTON, L. M., SEVERSON, R. K., TINKER, L. F., NORTH, K. E., BECKER, N., BENAVENTE, Y., BOFFETTA, P., BRENNAN, P., FORETOVA, L., MAYNADIE, M., STAINES, A., LIGHTFOOT, T., CROUCH, S., SMITH, A., ROMAN, E., DIVER, W. R., OFFIT, K., ZELENETZ, A., KLEIN, R. J., VILLANO, D. J., ZHENG, T., ZHANG, Y., HOLFORD, T. R., TURNER, J., SOUTHEY, M. C., CLAVEL, J., VIRTAMO, J., WEINSTEIN, S., RIBOLI, E., VINEIS, P., KAAKS, R., BOEING, H., TJONNELAND, A., ANGELUCCI, E., DI LOLLO, S., RAIS, M., DE VIVO, I., GIOVANNUCCI, E., KRAFT, P., HUANG, J., *et al.* 2016. Genetically predicted longer telomere length is associated with increased risk of B-cell lymphoma subtypes. *Hum Mol Genet*, 25, 1663-76.
- MACIEJOWSKI, J. & DE LANGE, T. 2017. Telomeres in cancer: tumour suppression and genome instability. *Nat Rev Mol Cell Biol*, 18, 175-186.
- MACIEJOWSKI, J., LI, Y., BOSCO, N., CAMPBELL, P. J. & DE LANGE, T. 2015. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell*, 163, 1641-54.
- MAIGA, A., LEMIEUX, S., PABST, C., LAVALLEE, V. P., BOUVIER, M., SAUVAGEAU, G. & HEBERT, J. 2016. Transcriptome analysis of G protein-coupled receptors in distinct genetic subgroups of acute myeloid leukemia: identification of potential disease-specific targets. *Blood Cancer J*, 6, e431.
- MAKISHIMA, H. & MACIEJEWSKI, J. P. 2011. Pathogenesis and consequences of uniparental disomy in cancer. *Clin Cancer Res*, 17, 3913-23.
- MALAVASI, F., DEAGLIO, S., DAMLE, R., CUTRONA, G., FERRARINI, M. & CHIORAZZI, N. 2011. CD38 and chronic lymphocytic leukemia: a decade later. *Blood*, 118, 3470-8.
- MALKOVA, A. & IRA, G. 2013. Break-induced replication: functions and molecular mechanism. *Curr Opin Genet Dev*, 23, 271-9.
- MANSOURI, L., GRABOWSKI, P., DEGERMAN, S., SVENSON, U., GUNNARSSON, R., CAHILL, N., SMEDBY, K. E., GEISLER, C., JULIUSSON, G., ROOS, G. & ROSENQUIST, R. 2013. Short telomere length is associated with NOTCH1/SF3B1/TP53 aberrations and poor outcome in newly diagnosed chronic lymphocytic leukemia patients. *Am J Hematol*, 88, 647-51.
- MAO, P., LIU, J., ZHANG, Z., ZHANG, H., LIU, H., GAO, S., RONG, Y. S. & ZHAO, Y. 2016. Homologous recombination-dependent repair of telomeric DSBs in proliferating human cells. *Nat Commun*, 7, 12154.
- MARTENS, U. M., ZIJLMANS, J. M., POON, S. S., DRAGOWSKA, W., YUI, J., CHAVEZ, E. A., WARD, R. K. & LANSDORP, P. M. 1998. Short telomeres on human chromosome 17p. *Nat Genet*, 18, 76-80.
- MARTINEZ-TRILLOS, A., QUESADA, V., VILLAMOR, N., PUENTE, X. S., LOPEZ-OTIN, C. & CAMPO, E. 2013. Recurrent gene mutations in CLL. *Adv Exp Med Biol*, 792, 87-107.
- MARTINEZ, P. & BLASCO, M. A. 2011. Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. *Nat Rev Cancer*, 11, 161-76.
- MARTINEZ, P., THANASOULA, M., MUNOZ, P., LIAO, C., TEJERA, A., MCNEES, C., FLORES, J. M., FERNANDEZ-CAPETILLO, O., TARSOUNAS, M. & BLASCO, M. A. 2009. Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. *Genes Dev*, 23, 2060-75.
- MARZEC, P., ARMENISE, C., PEROT, G., ROUMELIOTI, F. M., BASYUK, E., GAGOS, S., CHIBON, F. & DEJARDIN, J. 2015. Nuclear-receptor-mediated telomere insertion leads to genome instability in ALT cancers. *Cell*, 160, 913-27.
- MATEOS-GOMEZ, P. A., GONG, F., NAIR, N., MILLER, K. M., LAZZERINI-DENCHI, E. & SFEIR, A. 2015. Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature*, 518, 254-7.

- MATSUZAKI, H., FUJIMOTO, T., OTA, T., OGAWA, M., TSUNODA, T., DOI, K., HAMABASHIRI, M., TANAKA, M. & SHIRASAWA, S. 2012. Tespa1 is a novel inositol 1,4,5-trisphosphate receptor binding protein in T and B lymphocytes. *FEBS Open Bio*, 2, 255-9.
- MCCLINTOCK, B. 1941. The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics*, 26, 234-82.
- MCDOWELL, D. G., BURNS, N. A. & PARKES, H. C. 1998. Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. *Nucleic Acids Res*, 26, 3340-7.
- MCGRANAHAN, N. & SWANTON, C. 2017. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*, 168, 613-628.
- MCNEES, C. J., TEJERA, A. M., MARTINEZ, P., MURGA, M., MULERO, F., FERNANDEZ-CAPETILLO, O. & BLASCO, M. A. 2010. ATR suppresses telomere fragility and recombination but is dispensable for elongation of short telomeres by telomerase. *J Cell Biol*, 188, 639-52.
- MEEHAN, R. S. & CHEN, A. P. 2016. New treatment option for ovarian cancer: PARP inhibitors. *Gynecol Oncol Res Pract*, 3, 3.
- MEFFORD, H. C., LINARDOPOULOU, E., COIL, D., VAN DEN ENGH, G. & TRASK, B. J. 2001. Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes. *Hum Mol Genet*, 10, 2363-72.
- MEFFORD, H. C. & TRASK, B. J. 2002. The complex structure and dynamic evolution of human subtelomeres. *Nat Rev Genet*, 3, 91-102.
- 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 Res*, 37, 6225-38.
- MEYERSON, M., COUNTER, C. M., EATON, E. N., ELLISEN, L. W., STEINER, P., CADDLE, S. D., ZIAUGRA, L., BEIJERSBERGEN, R. L., DAVIDOFF, M. J., LIU, Q., BACCHETTI, S., HABER, D. A. & WEINBERG, R. A. 1997. hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. *Cell*, 90, 785-95.
- MILLER, D., REYNOLDS, G. E., MEJIA, R., STARK, J. M. & MURNANE, J. P. 2011. Subtelomeric regions in mammalian cells are deficient in DNA double-strand break repair. *DNA Repair (Amst)*, 10, 536-44.
- MIMITOU, E. P. & SYMINGTON, L. S. 2010. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO J*, 29, 3358-69.
- MOHR, S., DOEBELE, C., COMOGLIO, F., BERG, T., BECK, J., BOHNENBERGER, H., ALEXE, G., CORSO, J., STROBEL, P., WACHTER, A., BEISSBARTH, T., SCHNUTGEN, F., CREMER, A., HAETSCHER, N., GOLLNER, S., ROUHI, A., PALMQVIST, L., RIEGER, M. A., SCHROEDER, T., BONIG, H., MULLER-TIDOW, C., KUCHENBAUER, F., SCHUTZ, E., GREEN, A. R., URLAUB, H., STEGMAIER, K., HUMPHRIES, R. K., SERVE, H. & OELLERICH, T. 2017. Hoxa9 and Meis1 Cooperatively Induce Addiction to Syk Signaling by Suppressing miR-146a in Acute Myeloid Leukemia. *Cancer Cell*, 31, 549-562 e11.
- MONTERO, J. J., LOPEZ DE SILANES, I., GRANA, O. & BLASCO, M. A. 2016. Telomeric RNAs are essential to maintain telomeres. *Nat Commun*, 7, 12534.
- MOOTHA, V. K., LINDGREN, C. M., ERIKSSON, K. F., SUBRAMANIAN, A., SIHAG, S., LEHAR, J., PUIGSERVER, P., CARLSSON, E., RIDDERSTRALE, M., LAURILA, E., HOUSTIS, N., DALY, M. J., PATTERSON, N., MESIROV, J. P., GOLUB, T. R., TAMAYO, P., SPIEGELMAN, B.,

- LANDER, E. S., HIRSCHHORN, J. N., ALTSHULER, D. & GROOP, L. C. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*, 34, 267-73.
- MORIN, G. B. 1989. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell*, 59, 521-9.
- MOURIER, T., HANSEN, A. J., WILLERSLEV, E. & ARCTANDER, P. 2001. The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol Biol Evol*, 18, 1833-7.
- MOYNAHAN, M. E. & JASIN, M. 2010. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol*, 11, 196-207.
- MOYZIS, R. K., BUCKINGHAM, J. M., CRAM, L. S., DANI, M., DEAVEN, L. L., JONES, M. D., MEYNE, J., RATLIFF, R. L. & WU, J. R. 1988. A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A*, 85, 6622-6.
- MULLER, H. J. 1938. The Remaking of Chromosomes. *Collecting Net*, 13, 1181-1198.
- MUNIR, T., HOWARD, D. R., MCPARLAND, L., POCOOCK, C., RAWSTRON, A. C., HOCKADAY, A., VARGHESE, A., HAMBLIN, M., BLOOR, A., PETTITT, A., FEGAN, C., BLUNDELL, J., GRIBBEN, J. G., PHILLIPS, D. & HILLMEN, P. 2017. Results of the randomized phase IIB ADMIRE trial of FCR with or without mitoxantrone in previously untreated CLL. *Leukemia*.
- MURAKI, K., HAN, L., MILLER, D. & MURNANE, J. P. 2015. Processing by MRE11 is involved in the sensitivity of subtelomeric regions to DNA double-strand breaks. *Nucleic Acids Res*, 43, 7911-30.
- MURNANE, J. P. 2006. Telomeres and chromosome instability. *DNA Repair (Amst)*, 5, 1082-92.
- MURNANE, J. P. 2012. Telomere dysfunction and chromosome instability. *Mutat Res*, 730, 28-36.
- MURNANE, J. P., SABATIER, L., MARDER, B. A. & MORGAN, W. F. 1994. Telomere dynamics in an immortal human cell line. *EMBO J*, 13, 4953-62.
- NADEU, F., DELGADO, J., ROYO, C., BAUMANN, T., STANKOVIC, T., PINYOL, M., JARES, P., NAVARRO, A., MARTIN-GARCIA, D., BEA, S., SALAVERRIA, I., OLDREIVE, C., AYMERICH, M., SUAREZ-CISNEROS, H., ROZMAN, M., VILLAMOR, N., COLOMER, D., LOPEZ-GUILLERMO, A., GONZALEZ, M., ALCOCEBA, M., TEROL, M. J., COLADO, E., PUENTE, X. S., LOPEZ-OTIN, C., ENJUANES, A. & CAMPO, E. 2016. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood*, 127, 2122-30.
- NAGEL, I., SZCZEPANOWSKI, M., MARTIN-SUBERO, J. I., HARDER, L., AKASAKA, T., AMMERPOHL, O., CALLET-BAUCHU, E., GASCOYNE, R. D., GESK, S., HORSMAN, D., KLAPPER, W., MAJID, A., MARTINEZ-CLIMENT, J. A., STILGENBAUER, S., TONNIES, H., DYER, M. J. & SIEBERT, R. 2010. Deregulation of the telomerase reverse transcriptase (TERT) gene by chromosomal translocations in B-cell malignancies. *Blood*, 116, 1317-20.
- NANDAKUMAR, J., BELL, C. F., WEIDENFELD, I., ZAUG, A. J., LEINWAND, L. A. & CECH, T. R. 2012. The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature*, 492, 285-9.
- NANDAKUMAR, J. & CECH, T. R. 2013. Finding the end: recruitment of telomerase to telomeres. *Nat Rev Mol Cell Biol*, 14, 69-82.
- NEWMAN, E. A., LU, F., BASHLLARI, D., WANG, L., OPIPARI, A. W. & CASTLE, V. P. 2015. Alternative NHEJ Pathway Components Are Therapeutic Targets in High-Risk Neuroblastoma. *Mol Cancer Res*, 13, 470-82.

- NIKITIN, E. A., MALAKHO, S. G., BIDERMAN, B. V., BARANOVA, A. V., LORIE, Y. Y., SHEVELEV, A. Y., PEKLO, M. M., VLASIK, T. N., MOSKALEV, E. A., ZINGERMAN, B. V., VOROB'EV, I. A., POLTARAUS, A. B., SUDARIKOV, A. B. & VOROBEV, A. I. 2007. Expression level of lipoprotein lipase and dystrophin genes predict survival in B-cell chronic lymphocytic leukemia. *Leuk Lymphoma*, 48, 912-22.
- NORRBACK, K. F., HULTDIN, M., DAHLENBORG, K., OSTERMAN, P., CARLSSON, R. & ROOS, G. 2001. Telomerase regulation and telomere dynamics in germinal centers. *Eur J Haematol*, 67, 309-17.
- NOWELL, P. C. 1976. The clonal evolution of tumor cell populations. *Science*, 194, 23-8.
- O'CALLAGHAN, N. J. & FENECH, M. 2011. A quantitative PCR method for measuring absolute telomere length. *Biol Proced Online*, 13, 3.
- O'HAGAN, R. C., CHANG, S., MASER, R. S., MOHAN, R., ARTANDI, S. E., CHIN, L. & DEPINHO, R. A. 2002. Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell*, 2, 149-55.
- OHKI, R., TSURIMOTO, T. & ISHIKAWA, F. 2001. *In vitro* Reconstitution of the End Replication Problem. *Molecular and Cellular Biology*, 21, 5753-5766.
- OJHA, J., AYRES, J., SECRETO, C., TSCHUMPER, R., RABE, K., VAN DYKE, D., SLAGER, S., SHANAFELT, T., FONSECA, R., KAY, N. E. & BRAGGIO, E. 2015. Deep sequencing identifies genetic heterogeneity and recurrent convergent evolution in chronic lymphocytic leukemia. *Blood*, 125, 492-8.
- OJHA, J., CODD, V., NELSON, C. P., SAMANI, N. J., SMIRNOV, I. V., MADSEN, N. R., HANSEN, H. M., DE SMITH, A. J., BRACCI, P. M., WIENCKE, J. K., WRENSCH, M. R., WIEMELS, J. L., WALSH, K. M. & GROUP, E. C. T. 2016. Genetic Variation Associated with Longer Telomere Length Increases Risk of Chronic Lymphocytic Leukemia. *Cancer Epidemiol Biomarkers Prev*, 25, 1043-9.
- OLOVNIKOV, A. M. 1971. [Principle of marginotomy in template synthesis of polynucleotides]. *Dokl Akad Nauk SSSR*, 201, 1496-9.
- OLOVNIKOV, A. M. 1973. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J Theor Biol*, 41, 181-90.
- OSCIER, D., DEARDEN, C., EREN, E., FEGAN, C., FOLLOWS, G., HILLMEN, P., ILLIDGE, T., MATUTES, E., MILLIGAN, D. W., PETTITT, A., SCHUH, A., WIMPERIS, J. & BRITISH COMMITTEE FOR STANDARDS IN, H. 2012. Guidelines on the diagnosis, investigation and management of chronic lymphocytic leukaemia. *Br J Haematol*, 159, 541-64.
- OSCIER, D., WADE, R., DAVIS, Z., MORILLA, A., BEST, G., RICHARDS, S., ELSE, M., MATUTES, E., CATOVSKY, D. & CHRONIC LYMPHOCYTIC LEUKAEMIA WORKING GROUP, U. K. N. C. R. I. 2010. Prognostic factors identified three risk groups in the LRF CLL4 trial, independent of treatment allocation. *Haematologica*, 95, 1705-12.
- OUILLETTE, P., SAIYA-CORK, K., SEYMOUR, E., LI, C., SHEDDEN, K. & MALEK, S. N. 2013. Clonal evolution, genomic drivers, and effects of therapy in chronic lymphocytic leukemia. *Clin Cancer Res*, 19, 2893-904.
- PALM, W. & DE LANGE, T. 2008. How shelterin protects mammalian telomeres. *Annu Rev Genet*, 42, 301-34.
- PAMPALONA, J., FRIAS, C., GENESCA, A. & TUSELL, L. 2012. Progressive telomere dysfunction causes cytokinesis failure and leads to the accumulation of polyploid cells. *PLoS Genet*, 8, e1002679.
- PEI, J., JHANWAR, S. C. & TESTA, J. R. 2012. Chromothripsis in a Case of TP53-Deficient Chronic Lymphocytic Leukemia. *Leuk Res Rep*, 1, 4-6.
- PEIFER, M., HERTWIG, F., ROELS, F., DREIDAX, D., GARTLGRUBER, M., MENON, R., KRAMER, A., RONCAIOLI, J. L., SAND, F., HEUCKMANN, J. M., IKRAM, F., SCHMIDT, R.,

- ACKERMANN, S., ENGESSER, A., KAHLERT, Y., VOGEL, W., ALTMULLER, J., NURNBERG, P., THIERRY-MIEG, J., THIERRY-MIEG, D., MARIAPPAN, A., HEYNCK, S., MARIOTTI, E., HENRICH, K. O., GLOECKNER, C., BOSCO, G., LEUSCHNER, I., SCHWEIGER, M. R., SAVELYEVA, L., WATKINS, S. C., SHAO, C., BELL, E., HOFER, T., ACHTER, V., LANG, U., THEISSEN, J., VOLLAND, R., SAADATI, M., EGGERT, A., DE WILDE, B., BERTHOLD, F., PENG, Z., ZHAO, C., SHI, L., ORTMANN, M., BUTTNER, R., PERNER, S., HERO, B., SCHRAMM, A., SCHULTE, J. H., HERRMANN, C., O'SULLIVAN, R. J., WESTERMANN, F., THOMAS, R. K. & FISCHER, M. 2015. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature*, 526, 700-4.
- PENA-DIAZ, J., BREGENHORN, S., GHODGAONKAR, M., FOLLONIER, C., ARTOLA-BORAN, M., CASTOR, D., LOPES, M., SARTORI, A. A. & JIRICNY, J. 2012. Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol Cell*, 47, 669-80.
- PEPPER, C., BAIRD, D. & FEGAN, C. 2014. Telomere analysis to predict chronic lymphocytic leukemia outcome: a STELA test to change clinical practice? *Expert Rev Hematol*, 7, 701-3.
- PEPPER, C., WARD, R., LIN, T. T., BRENNAN, P., STARCZYNSKI, J., MUSSON, M., ROWNTREE, C., BENTLEY, P., MILLS, K., PRATT, G. & FEGAN, C. 2007. Highly purified CD38+ and CD38- sub-clones derived from the same chronic lymphocytic leukemia patient have distinct gene expression signatures despite their monoclonal origin. *Leukemia*, 21, 687-96.
- PETTITT, A. R. 2003. Mechanism of action of purine analogues in chronic lymphocytic leukaemia. *Br J Haematol*, 121, 692-702.
- PETTITT, A. R., SHERRINGTON, P. D., STEWART, G., CAWLEY, J. C., TAYLOR, A. M. & STANKOVIC, T. 2001. p53 dysfunction in B-cell chronic lymphocytic leukemia: inactivation of ATM as an alternative to TP53 mutation. *Blood*, 98, 814-22.
- PICKETT, H. A. & REDDEL, R. R. 2015. Molecular mechanisms of activity and derepression of alternative lengthening of telomeres. *Nat Struct Mol Biol*, 22, 875-80.
- PONCET, D., BELLEVILLE, A., T'KINT DE ROODENBEKE, C., ROBOREL DE CLIMENS, A., BEN SIMON, E., MERLE-BERAL, H., CALLET-BAUCHU, E., SALLES, G., SABATIER, L., DELIC, J. & GILSON, E. 2008. Changes in the expression of telomere maintenance genes suggest global telomere dysfunction in B-chronic lymphocytic leukemia. *Blood*, 111, 2388-91.
- PUENTE, X. S., BEA, S., VALDES-MAS, R., VILLAMOR, N., GUTIERREZ-ABRIL, J., MARTIN-SUBERO, J. I., MUNAR, M., RUBIO-PEREZ, C., JARES, P., AYMERICH, M., BAUMANN, T., BEEKMAN, R., BELVER, L., CARRIO, A., CASTELLANO, G., CLOT, G., COLADO, E., COLOMER, D., COSTA, D., DELGADO, J., ENJUANES, A., ESTIVILL, X., FERRANDO, A. A., GELPI, J. L., GONZALEZ, B., GONZALEZ, S., GONZALEZ, M., GUT, M., HERNANDEZ-RIVAS, J. M., LOPEZ-GUERRA, M., MARTIN-GARCIA, D., NAVARRO, A., NICOLAS, P., OROZCO, M., PAYER, A. R., PINYOL, M., PISANO, D. G., PUENTE, D. A., QUEIROS, A. C., QUESADA, V., ROMEO-CASABONA, C. M., ROYO, C., ROYO, R., ROZMAN, M., RUSSINOL, N., SALAVERRIA, I., STAMATOPOULOS, K., STUNNENBERG, H. G., TAMBORERO, D., TEROL, M. J., VALENCIA, A., LOPEZ-BIGAS, N., TORRENTS, D., GUT, I., LOPEZ-GUILLERMO, A., LOPEZ-OTIN, C. & CAMPO, E. 2015. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*, 526, 519-24.
- PUENTE, X. S. & LOPEZ-OTIN, C. 2013. The evolutionary biography of chronic lymphocytic leukemia. *Nat Genet*, 45, 229-31.
- PUENTE, X. S., PINYOL, M., QUESADA, V., CONDE, L., ORDONEZ, G. R., VILLAMOR, N., ESCARAMIS, G., JARES, P., BEA, S., GONZALEZ-DIAZ, M., BASSAGANYAS, L., BAUMANN, T., JUAN, M., LOPEZ-GUERRA, M., COLOMER, D., TUBIO, J. M., LOPEZ, C., NAVARRO, A., TORNADOR, C., AYMERICH, M., ROZMAN, M., HERNANDEZ, J. M., PUENTE, D. A., FREIJE, J. M., VELASCO, G., GUTIERREZ-FERNANDEZ, A., COSTA, D.,

- CARRIO, A., GUIJARRO, S., ENJUANES, A., HERNANDEZ, L., YAGUE, J., NICOLAS, P., ROMEO-CASABONA, C. M., HIMMELBAUER, H., CASTILLO, E., DOHM, J. C., DE SANJOSE, S., PIRIS, M. A., DE ALAVA, E., SAN MIGUEL, J., ROYO, R., GELPI, J. L., TORRENTS, D., OROZCO, M., PISANO, D. G., VALENCIA, A., GUIGO, R., BAYES, M., HEATH, S., GUT, M., KLATT, P., MARSHALL, J., RAINE, K., STEBBINGS, L. A., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J., GUT, I., LOPEZ-GUILLERMO, A., ESTIVILL, X., MONTSERRAT, E., LOPEZ-OTIN, C. & CAMPO, E. 2011. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*, 475, 101-5.
- PUIGGROS, A., DELGADO, J., RODRIGUEZ-VICENTE, A., COLLADO, R., AVENTIN, A., LUNO, E., GRAU, J., HERNANDEZ, J. A., MARUGAN, I., ARDANAZ, M., GONZALEZ, T., VALIENTE, A., OSMÁ, M., CALASANZ, M. J., SANZO, C., CARRIO, A., ORTEGA, M., SANTACRUZ, R., ABRISQUETA, P., ABELLA, E., BOSCH, F., CARBONELL, F., SOLE, F., HERNANDEZ, J. M., ESPINET, B., GRUPO COOPERATIVO ESPANOL DE CITOGENETICA, H. & GRUPO ESPANOL DE LEUCEMIA LINFATICA, C. 2013. Biallelic losses of 13q do not confer a poorer outcome in chronic lymphocytic leukaemia: analysis of 627 patients with isolated 13q deletion. *Br J Haematol*, 163, 47-54.
- PUIGGROS, A., VENTURAS, M., SALIDO, M., BLANCO, G., FERNANDEZ-RODRIGUEZ, C., COLLADO, R., VALIENTE, A., RUIZ-XIVILLE, N., CARRIO, A., ORTUNO, F. J., LUNO, E., CALASANZ, M. J., ARDANAZ, M. T., PINAN, M. A., TALAVERA, E., GONZALEZ, M. T., ORTEGA, M., MARUGAN, I., FERRER, A., GIMENO, E., BELLOSILLO, B., DELGADO, J., HERNANDEZ, J. A., HERNANDEZ-RIVAS, J. M., ESPINET, B., GRUPO COOPERATIVO ESPANOL DE CITOGENETICA, H. & GRUPO ESPANOL DE LEUCEMIA LINFATICA, C. 2014. Interstitial 13q14 deletions detected in the karyotype and translocations with concomitant deletion at 13q14 in chronic lymphocytic leukemia: different genetic mechanisms but equivalent poorer clinical outcome. *Genes Chromosomes Cancer*, 53, 788-97.
- QUEIROS, A. C., VILLAMOR, N., CLOT, G., MARTINEZ-TRILLOS, A., KULIS, M., NAVARRO, A., PENAS, E. M., JAYNE, S., MAJID, A., RICHTER, J., BERGMANN, A. K., KOLAROVA, J., ROYO, C., RUSSINOL, N., CASTELLANO, G., PINYOL, M., BEA, S., SALAVERRIA, I., LOPEZ-GUERRA, M., COLOMER, D., AYMERICH, M., ROZMAN, M., DELGADO, J., GINE, E., GONZALEZ-DIAZ, M., PUENTE, X. S., SIEBERT, R., DYER, M. J., LOPEZ-OTIN, C., ROZMAN, C., CAMPO, E., LOPEZ-GUILLERMO, A. & MARTIN-SUBERO, J. I. 2015. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia*, 29, 598-605.
- QUESADA, V., CONDE, L., VILLAMOR, N., ORDONEZ, G. R., JARES, P., BASSAGANYAS, L., RAMSAY, A. J., BEA, S., PINYOL, M., MARTINEZ-TRILLOS, A., LOPEZ-GUERRA, M., COLOMER, D., NAVARRO, A., BAUMANN, T., AYMERICH, M., ROZMAN, M., DELGADO, J., GINE, E., HERNANDEZ, J. M., GONZALEZ-DIAZ, M., PUENTE, D. A., VELASCO, G., FREIJE, J. M., TUBIO, J. M., ROYO, R., GELPI, J. L., OROZCO, M., PISANO, D. G., ZAMORA, J., VAZQUEZ, M., VALENCIA, A., HIMMELBAUER, H., BAYES, M., HEATH, S., GUT, M., GUT, I., ESTIVILL, X., LOPEZ-GUILLERMO, A., PUENTE, X. S., CAMPO, E. & LOPEZ-OTIN, C. 2012. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*, 44, 47-52.
- RAI, K. R., SAWITSKY, A., CRONKITE, E. P., CHANANA, A. D., LEVY, R. N. & PASTERNAK, B. S. 1975. Clinical staging of chronic lymphocytic leukemia. *Blood*, 46, 219-34.
- RAI, R., ZHENG, H., HE, H., LUO, Y., MULTANI, A., CARPENTER, P. B. & CHANG, S. 2010. The function of classical and alternative non-homologous end-joining pathways in the fusion of dysfunctional telomeres. *EMBO J*, 29, 2598-610.

- RAJASAGI, M., SHUKLA, S. A., FRITSCH, E. F., KESKIN, D. B., DELUCA, D., CARMONA, E., ZHANG, W., SOUGNEZ, C., CIBULSKIS, K., SIDNEY, J., STEVENSON, K., RITZ, J., NEUBERG, D., BRUSIC, V., GABRIEL, S., LANDER, E. S., GETZ, G., HACOEN, N. & WU, C. J. 2014. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood*, 124, 453-62.
- RAMSAY, A. J., QUESADA, V., FORONDA, M., CONDE, L., MARTINEZ-TRILLOS, A., VILLAMOR, N., RODRIGUEZ, D., KWARCIAK, A., GARABAYA, C., GALLARDO, M., LOPEZ-GUERRA, M., LOPEZ-GUILLERMO, A., PUENTE, X. S., BLASCO, M. A., CAMPO, E. & LOPEZ-OTIN, C. 2013. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat Genet*, 45, 526-30.
- REBHANDL, S., HUEMER, M., GASSNER, F. J., ZABORSKY, N., HEBENSTREIT, D., CATAKOVIC, K., GROSSINGER, E. M., GREIL, R. & GEISBERGER, R. 2014. APOBEC3 signature mutations in chronic lymphocytic leukemia. *Leukemia*, 28, 1929-32.
- REDDEL, R. R. 2000. The role of senescence and immortalization in carcinogenesis. *Carcinogenesis*, 21, 477-84.
- REDDEL, R. R., BRYAN, T. M., COLGIN, L. M., PERREM, K. T. & YEAGER, T. R. 2001. Alternative Lengthening of Telomeres in Human Cells. *Radiation Research*, 155, 194-200.
- RIBALLO, E., KUHNE, M., RIEF, N., DOHERTY, A., SMITH, G. C., RECIO, M. J., REIS, C., DAHM, K., FRICKE, A., KREMLER, A., PARKER, A. R., JACKSON, S. P., GENNERY, A., JEGGO, P. A. & LOBRICH, M. 2004. A pathway of double-strand break rejoining dependent upon ATM, Artemis, and proteins locating to gamma-H2AX foci. *Mol Cell*, 16, 715-24.
- RIBONI, R., CASATI, A., NARDO, T., ZACCARO, E., FERRETTI, L., NUZZO, F. & MONDELLO, C. 1997. Telomeric fusions in cultured human fibroblasts as a source of genomic instability. *Cancer Genet Cytogenet*, 95, 130-6.
- RICCA, I., ROCCI, A., DRANDI, D., FRANCESE, R., COMPAGNO, M., LOBETTI BODONI, C., DE MARCO, F., ASTOLFI, M., MONITILLO, L., VALLET, S., CALVI, R., FICARA, F., OMEDE, P., ROSATO, R., GALLAMINI, A., MARINONE, C., BERGUI, L., BOCCADORO, M., TARELLA, C. & LADETTO, M. 2007. Telomere length identifies two different prognostic subgroups among VH-unmutated B-cell chronic lymphocytic leukemia patients. *Leukemia*, 21, 697-705.
- RICE, C., SHASTRULA, P. K., KOSENKOV, A. V., HILLS, R., BAIRD, D. M., SHOWE, L. C., DOUKOV, T., JANICKI, S. & SKORDALAKES, E. 2017. Structural and functional analysis of the human POT1-TPP1 telomeric complex. *Nat Commun*, 8, 14928.
- RICHARDS, E. J. & AUSUBEL, F. M. 1988. Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell*, 53, 127-36.
- RICHARDSON, C. J., BROENSTRUP, M., FINGAR, D. C., JULICH, K., BALLIF, B. A., GYGI, S. & BLENIS, J. 2004. SKAR is a specific target of S6 kinase 1 in cell growth control. *Curr Biol*, 14, 1540-9.
- RIETHMAN, H., AMBROSINI, A., CASTANEDA, C., FINKLESTEIN, J., HU, X. L., MUDUNURI, U., PAUL, S. & WEI, J. 2004. Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Res*, 14, 18-28.
- RIETHMAN, H., AMBROSINI, A. & PAUL, S. 2005. Human subtelomere structure and variation. *Chromosome Res*, 13, 505-15.
- ROBAK, T., DMOSZYNSKA, A., SOLAL-CELINEY, P., WARZOCHA, K., LOSCERTALES, J., CATALANO, J., AFANASIEV, B. V., LARRATT, L., GEISLER, C. H., MONTILLO, M., ZYUZGIN, I., GANLY, P. S., DARTIGEAS, C., ROSTA, A., MAURER, J., MENDILA, M., SAVILLE, M. W., VALENTE, N., WENGER, M. K. & MOISEEV, S. I. 2010. Rituximab plus fludarabine and cyclophosphamide prolongs progression-free survival compared

- with fludarabine and cyclophosphamide alone in previously treated chronic lymphocytic leukemia. *J Clin Oncol*, 28, 1756-65.
- ROBERT, I., DANTZER, F. & REINA-SAN-MARTIN, B. 2009. Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. *J Exp Med*, 206, 1047-56.
- ROBIN, J. D., LUDLOW, A. T., BATTEN, K., MAGDINIER, F., STADLER, G., WAGNER, K. R., SHAY, J. W. & WRIGHT, W. E. 2014. Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. *Genes Dev*, 28, 2464-76.
- ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nat Biotechnol*, 29, 24-6.
- RODRIGUEZ-PAREDES, M. & ESTELLER, M. 2011. Cancer epigenetics reaches mainstream oncology. *Nat Med*, 17, 330-9.
- ROGER, L., JONES, R. E., HEPPEL, N. H., WILLIAMS, G. T., SAMPSON, J. R. & BAIRD, D. M. 2013. Extensive telomere erosion in the initiation of colorectal adenomas and its association with chromosomal instability. *J Natl Cancer Inst*, 105, 1202-11.
- ROOS, G., KROBER, A., GRABOWSKI, P., KIENLE, D., BUHLER, A., DOHNER, H., ROSENQUIST, R. & STILGENBAUER, S. 2008. Short telomeres are associated with genetic complexity, high-risk genomic aberrations, and short survival in chronic lymphocytic leukemia. *Blood*, 111, 2246-52.
- ROSSI, D., LOBETTI BODONI, C., GENUARDI, E., MONITILLO, L., DRANDI, D., CERRI, M., DEAMBROGI, C., RICCA, I., ROCCI, A., FERRERO, S., BERNOCCO, E., CAPELLO, D., DE PAOLI, L., BERGUI, L., BOI, M., OMEDE, P., MASSAIA, M., TARELLA, C., PASSERA, R., BOCCADORO, M., GAIDANO, G. & LADETTO, M. 2009. Telomere length is an independent predictor of survival, treatment requirement and Richter's syndrome transformation in chronic lymphocytic leukemia. *Leukemia*, 23, 1062-72.
- ROY-ENGEL, A. M., CARROLL, M. L., VOGEL, E., GARBER, R. K., NGUYEN, S. V., SALEM, A. H., BATZER, M. A. & DEININGER, P. L. 2001. Alu insertion polymorphisms for the study of human genomic diversity. *Genetics*, 159, 279-90.
- ROY, K., WU, Y., MEITZLER, J. L., JUHASZ, A., LIU, H., JIANG, G., LU, J., ANTONY, S. & DOROSHOW, J. H. 2015. NADPH oxidases and cancer. *Clin Sci (Lond)*, 128, 863-75.
- ROZEN, S. & SKALETSKY, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 132, 365-86.
- ROZOVSKI, U., HAZAN-HALEVY, I., BARZILAI, M., KEATING, M. J. & ESTROV, Z. 2016. Metabolism pathways in chronic lymphocytic leukemia. *Leuk Lymphoma*, 57, 758-65.
- RUDD, M. K., FRIEDMAN, C., PARGHI, S. S., LINARDOPOULOU, E. V., HSU, L. & TRASK, B. J. 2007. Elevated rates of sister chromatid exchange at chromosome ends. *PLoS Genet*, 3, e32.
- RUFER, N., DRAGOWSKA, W., THORNBURY, G., ROOSNEK, E. & LANSDORP, P. M. 1998. Telomere length dynamics in human lymphocyte subpopulations measured by flow cytometry. *Nat Biotechnol*, 16, 743-7.
- SAHIN, E., COLLA, S., LIESA, M., MOSLEHI, J., MULLER, F. L., GUO, M., COOPER, M., KOTTON, D., FABIAN, A. J., WALKEY, C., MASER, R. S., TONON, G., FOERSTER, F., XIONG, R., WANG, Y. A., SHUKLA, S. A., JASKELIOFF, M., MARTIN, E. S., HEFFERNAN, T. P., PROTOPOPOV, A., IVANOVA, E., MAHONEY, J. E., KOST-ALIMOVA, M., PERRY, S. R., BRONSON, R., LIAO, R., MULLIGAN, R., SHIRIHAI, O. S., CHIN, L. & DEPINHO, R. A. 2011. Telomere dysfunction induces metabolic and mitochondrial compromise. *Nature*, 470, 359-65.

- SAKOFSKY, C. J., AYYAR, S., DEEM, A. K., CHUNG, W. H., IRA, G. & MALKOVA, A. 2015. Translesion Polymerases Drive Microhomology-Mediated Break-Induced Replication Leading to Complex Chromosomal Rearrangements. *Mol Cell*, 60, 860-72.
- SALAVERRIA, I., MARTIN-GARCIA, D., LOPEZ, C., CLOT, G., GARCIA-ARAGONES, M., NAVARRO, A., DELGADO, J., BAUMANN, T., PINYOL, M., MARTIN-GUERRERO, I., CARRIO, A., COSTA, D., QUEIROS, A. C., JAYNE, S., AYMERICH, M., VILLAMOR, N., COLOMER, D., GONZALEZ, M., LOPEZ-GUILLERMO, A., CAMPO, E., DYER, M. J., SIEBERT, R., ARMENGOL, L. & BEA, S. 2015. Detection of chromothripsis-like patterns with a custom array platform for chronic lymphocytic leukemia. *Genes Chromosomes Cancer*, 54, 668-80.
- SAMBROOK, J., FRITSCH, E.F. AND MANIATIS, T. 1989. *Molecular cloning: a Laboratory Manual*, New York, Cold Spring Harbor Laboratory Press.
- SANDIN, S. & RHODES, D. 2014. Telomerase structure. *Curr Opin Struct Biol*, 25, 104-10.
- SAREK, G., MARZEC, P., MARGALEF, P. & BOULTON, S. J. 2015. Molecular basis of telomere dysfunction in human genetic diseases. *Nat Struct Mol Biol*, 22, 867-74.
- SCHILLING, G., PENAS, E. M., JANJETOVIC, S., OLIVEIRA-FERRER, L., BRAIG, M., BEHRMANN, P., BOKEMEYER, C. & DIERLAMM, J. 2013. Molecular characterization of chromosomal band 5p15.33: a recurrent breakpoint region in mantle cell lymphoma involving the TERT-CLPTM1L locus. *Leuk Res*, 37, 280-6.
- SEIFERT, M., SELLMANN, L., BLOEHDORN, J., WEIN, F., STILGENBAUER, S., DURIG, J. & KUPPERS, R. 2012. Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J Exp Med*, 209, 2183-98.
- SELLMANN, L., SCHOLTYSIK, R., DE BEER, D., EISELE, L., KLEIN-HITPASS, L., NUCKEL, H., DUHRSEN, U., DURIG, J., ROTH, A. & BAERLOCHER, G. M. 2016. Shorter telomeres correlate with an increase in the number of uniparental disomies in patients with chronic lymphocytic leukemia. *Leuk Lymphoma*, 57, 590-5.
- SFEIR, A. & DE LANGE, T. 2012. Removal of shelterin reveals the telomere end-protection problem. *Science*, 336, 593-7.
- SFEIR, A., KABIR, S., VAN OVERBEEK, M., CELLI, G. B. & DE LANGE, T. 2010. Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science*, 327, 1657-61.
- SHAKIROV, E. V. & SHIPPEN, D. E. 2004. Length regulation and dynamics of individual telomere tracts in wild-type Arabidopsis. *Plant Cell*, 16, 1959-67.
- SHAY, J. W., PEREIRA-SMITH, O. M. & WRIGHT, W. E. 1991. A role for both RB and p53 in the regulation of human cellular senescence. *Exp Cell Res*, 196, 33-9.
- SIMPSON, K., JONES, R. E., GRIMSTEAD, J. W., HILLS, R., PEPPER, C. & BAIRD, D. M. 2015. Telomere fusion threshold identifies a poor prognostic subset of breast cancer patients. *Mol Oncol*, 9, 1186-93.
- SIMSEK, D., BRUNET, E., WONG, S. Y., KATYAL, S., GAO, Y., MCKINNON, P. J., LOU, J., ZHANG, L., LI, J., REBAR, E. J., GREGORY, P. D., HOLMES, M. C. & JASIN, M. 2011. DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS Genet*, 7, e1002080.
- SMITH, J. S., CHEN, Q., YATSUNYK, L. A., NICOLUDIS, J. M., GARCIA, M. S., KRANASTER, R., BALASUBRAMANIAN, S., MONCHAUD, D., TEULADE-FICHO, M. P., ABRAMOWITZ, L., SCHULTZ, D. C. & JOHNSON, F. B. 2011. Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat Struct Mol Biol*, 18, 478-85.
- SMITH, S. & DE LANGE, T. 2000. Tankyrase promotes telomere elongation in human cells. *Curr Biol*, 10, 1299-302.

- SMOGORZEWSKA, A., KARLSEDER, J., HOLTGREVE-GREZ, H., JAUCH, A. & DE LANGE, T. 2002. DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Curr Biol*, 12, 1635-44.
- SPIER, C. M., KJELDSBERG, C. R., HEAD, D. R., DIFIORE, K. C. & TUDOR, B. 1985. Chronic lymphocytic leukemia in young adults. *Am J Clin Pathol*, 84, 675-8.
- SRINIVASAINAGENDRA, V., SANDEL, M. W., SINGH, B., SUNDARESAN, A., MOOGA, V. P., BAJPAI, P., TIWARI, H. K. & SINGH, K. K. 2017. Migration of mitochondrial DNA in the nuclear genome of colorectal adenocarcinoma. *Genome Med*, 9, 31.
- STACCHINI, A., ARAGNO, M., VALLARIO, A., ALFARANO, A., CIRCOSTA, P., GOTTARDI, D., FALDELLA, A., REGE-CAMBRIN, G., THUNBERG, U., NILSSON, K. & CALIGARIS-CAPPIO, F. 1999. MEC1 and MEC2: two new cell lines derived from B-chronic lymphocytic leukaemia in polymphocytoid transformation. *Leuk Res*, 23, 127-36.
- STAMATOPOULOS, B., TIMBS, A., BRUCE, D., SMITH, T., CLIFFORD, R., ROBBE, P., BURNS, A., VAVOULIS, D. V., LOPEZ, L., ANTONIOU, P., MASON, J., DREAU, H. & SCHUH, A. 2017. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. *Leukemia*, 31, 837-845.
- STANKIEWICZ, P. 2016. One pedigree we all may have come from - did Adam and Eve have the chromosome 2 fusion? *Mol Cytogenet*, 9, 72.
- STEPHENS, K., WEAVER, M., LEPPIG, K. A., MARUYAMA, K., EMANUEL, P. D., LE BEAU, M. M. & SHANNON, K. M. 2006. Interstitial uniparental isodisomy at clustered breakpoint intervals is a frequent mechanism of NF1 inactivation in myeloid malignancies. *Blood*, 108, 1684-9.
- STEPHENS, P. J., GREENMAN, C. D., FU, B., YANG, F., BIGNELL, G. R., MUDIE, L. J., PLEASANCE, E. D., LAU, K. W., BEARE, D., STEBBINGS, L. A., MCLAREN, S., LIN, M. L., MCBRIDE, D. J., VARELA, I., NIK-ZAINAL, S., LEROY, C., JIA, M., MENZIES, A., BUTLER, A. P., TEAGUE, J. W., QUAIL, M. A., BURTON, J., SWERDLOW, H., CARTER, N. P., MORSBERGER, L. A., IACOBUIO-DONAHUE, C., FOLLOWS, G. A., GREEN, A. R., FLANAGAN, A. M., STRATTON, M. R., FUTREAL, P. A. & CAMPBELL, P. J. 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144, 27-40.
- STERN, J. L., ZYNER, K. G., PICKETT, H. A., COHEN, S. B. & BRYAN, T. M. 2012. Telomerase recruitment requires both TCAB1 and Cajal bodies independently. *Mol Cell Biol*, 32, 2384-95.
- STILGENBAUER, S., BULLINGER, L., BENNER, A., WILDENBERGER, K., BENTZ, M., DOHNER, K., HO, A. D., LICHTER, P. & DOHNER, H. 1999. Incidence and clinical significance of 6q deletions in B cell chronic lymphocytic leukemia. *Leukemia*, 13, 1331-4.
- STOHR, B. A., XU, L. & BLACKBURN, E. H. 2010. The terminal telomeric DNA sequence determines the mechanism of dysfunctional telomere fusion. *Mol Cell*, 39, 307-14.
- STONG, N., DENG, Z., GUPTA, R., HU, S., PAUL, S., WEINER, A. K., EICHLER, E. E., GRAVES, T., FRONICK, C. C., COURTNEY, L., WILSON, R. K., LIEBERMAN, P. M., DAVULURI, R. V. & RIETHMAN, H. 2014. Subtelomeric CTCF and cohesin binding site organization using improved subtelomere assemblies and a novel annotation pipeline. *Genome Res*, 24, 1039-50.
- STRATI, P. & SHANAFELT, T. D. 2015. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood*, 126, 454-62.
- STRATTON, M. R., CAMPBELL, P. J. & FUTREAL, P. A. 2009. The cancer genome. *Nature*, 458, 719-24.
- STREFFORD, J. C., KADALAYIL, L., FORSTER, J., ROSE-ZERILLI, M. J., PARKER, A., LIN, T. T., HEPPEL, N., NORRIS, K., GARDINER, A., DAVIES, Z., GONZALEZ DE CASTRO, D., ELSE, M., STEELE, A. J., PARKER, H., STANKOVIC, T., PEPPER, C., FEGAN, C., BAIRD, D.,

- COLLINS, A., CATOVSKY, D. & OSCIER, D. G. 2015. Telomere length predicts progression and overall survival in chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Leukemia*, 29, 2411-4.
- STRUSKI, S., HELIAS, C., GERVAIS, C., AUDHUY, B., ZAMFIR, A., HERBRECHT, R. & LESSARD, M. 2007. 13q deletions in B-cell lymphoproliferative disorders: frequent association with translocation. *Cancer Genet Cytogenet*, 174, 151-60.
- SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102, 15545-50.
- SUNG, P. 1994. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science*, 265, 1241-3.
- SUNG, P., KREJCI, L., VAN KOMEN, S. & SEHORN, M. G. 2003. Rad51 recombinase and recombination mediators. *J Biol Chem*, 278, 42729-32.
- SYMINGTON, L. S. 2016. Mechanism and regulation of DNA end resection in eukaryotes. *Crit Rev Biochem Mol Biol*, 51, 195-212.
- TAM, C. S. & STILGENBAUER, S. 2015. How best to manage patients with chronic lymphocytic leukemia with 17p deletion and/or TP53 mutation? *Leuk Lymphoma*, 56, 587-93.
- TEJERA, A. M., STAGNO D'ALCONTRES, M., THANASOULA, M., MARION, R. M., MARTINEZ, P., LIAO, C., FLORES, J. M., TARSOUNAS, M. & BLASCO, M. A. 2010. TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. *Dev Cell*, 18, 775-89.
- THANASOULA, M., ESCANDELL, J. M., SUWAKI, N. & TARSOUNAS, M. 2012. ATM/ATR checkpoint activation downregulates CDC25C to prevent mitotic entry with uncapped telomeres. *EMBO J*, 31, 3398-410.
- TOKUTAKE, Y., MATSUMOTO, T., WATANABE, T., MAEDA, S., TAHARA, H., SAKAMOTO, S., NIIDA, H., SUGIMOTO, M., IDE, T. & FURUICHI, Y. 1998. Extra-chromosomal telomere repeat DNA in telomerase-negative immortalized cell lines. *Biochem Biophys Res Commun*, 247, 765-72.
- TOMIDA, J., TAKATA, K., LANGE, S. S., SCHIBLER, A. C., YOUSEFZADEH, M. J., BHETAWAL, S., DENT, S. Y. & WOOD, R. D. 2015. REV7 is essential for DNA damage tolerance via two REV3L binding sites in mammalian DNA polymerase zeta. *Nucleic Acids Res*, 43, 1000-11.
- TOMIMATSU, N., MUKHERJEE, B., CATHERINE HARDEBECK, M., ILCHEVA, M., VANESSA CAMACHO, C., LOUISE HARRIS, J., PORTEUS, M., LLORENTE, B., KHANNA, K. K. & BURMA, S. 2014. Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice. *Nat Commun*, 5, 3561.
- 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 Rep*, 13, 1633-46.
- TRASK, B. J., FRIEDMAN, C., MARTIN-GALLARDO, A., ROWEN, L., AKINBAMI, C., BLANKENSHIP, J., COLLINS, C., GIORGI, D., IADONATO, S., JOHNSON, F., KUO, W. L., MASSA, H., MORRISH, T., NAYLOR, S., NGUYEN, O. T., ROUQUIER, S., SMITH, T., WONG, D. J., YOUNGBLOM, J. & VAN DEN ENGH, G. 1998. Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet*, 7, 13-26.
- TSIMBERIDOU, A. M., TAM, C., ABRUZZO, L. V., O'BRIEN, S., WIERDA, W. G., LERNER, S., KANTARJIAN, H. M. & KEATING, M. J. 2009. Chemoimmunotherapy may overcome the adverse prognostic significance of 11q deletion in previously untreated patients with chronic lymphocytic leukemia. *Cancer*, 115, 373-80.

- TURTLE, C. J., HAY, K. A., HANAFI, L. A., LI, D., CHERIAN, S., CHEN, X., WOOD, B., LOZANSKI, A., BYRD, J. C., HEIMFELD, S., RIDDELL, S. R. & MALONEY, D. G. 2017. Durable Molecular Remissions in Chronic Lymphocytic Leukemia Treated With CD19-Specific Chimeric Antigen Receptor-Modified T Cells After Failure of Ibrutinib. *J Clin Oncol*, 35, 3010-3020.
- TUTTON, S., AZZAM, G. A., STONG, N., VLADIMIROVA, O., WIEDMER, A., MONTEITH, J. A., BEISHLINE, K., WANG, Z., DENG, Z., RIETHMAN, H., MCMAHON, S. B., MURPHY, M. & LIEBERMAN, P. M. 2016. Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J*, 35, 193-207.
- VALENTIJN, L. J., KOSTER, J., ZWIJNENBURG, D. A., HASSELT, N. E., VAN SLUIS, P., VOLCKMANN, R., VAN NOESEL, M. M., GEORGE, R. E., TYTGAT, G. A., MOLENAAR, J. J. & VERSTEEG, R. 2015. TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat Genet*, 47, 1411-4.
- VAN STEENSEL, B. & DE LANGE, T. 1997. Control of telomere length by the human telomeric protein TRF1. *Nature*, 385, 740-3.
- VARGOVA, K., CURIK, N., BURDA, P., BASOVA, P., KULVAIT, V., POSPISIL, V., SAVVULIDI, F., KOKAVEC, J., NECAS, E., BERKOVA, A., OBRTLIKOVA, P., KARBAN, J., MRAZ, M., POSPISILOVA, S., MAYER, J., TRNENY, M., ZAVADIL, J. & STOPKA, T. 2011. MYB transcriptionally regulates the miR-155 host gene in chronic lymphocytic leukemia. *Blood*, 117, 3816-25.
- VETTERMANN, C. & SCHLISSEL, M. S. 2010. Allelic exclusion of immunoglobulin genes: models and mechanisms. *Immunol Rev*, 237, 22-42.
- WALSH, S. H., GRABOWSKI, P., BERGLUND, M., THUNBERG, U., THORSELIUS, M., TOBIN, G., ALESKOG, A., KARLSSON, K., SUNDSTROM, C., LAURELL, A., ENBLAD, G., ROSENQUIST, R. & ROOS, G. 2007. Telomere length and correlation with histopathogenesis in B-cell leukemias/lymphomas. *Eur J Haematol*, 78, 283-9.
- WANG, H., LIANG, L., FANG, J. Y. & XU, J. 2016. Somatic gene copy number alterations in colorectal cancer: new quest for cancer drivers and biomarkers. *Oncogene*, 35, 2011-9.
- WANG, L. E. A. 2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* 365.
- WANG, M., WU, W., WU, W., ROSIDI, B., ZHANG, L., WANG, H. & ILIAKIS, G. 2006. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res*, 34, 6170-82.
- WANG, Y., HUANG, J. W., CALSES, P., KEMP, C. J. & TANIGUCHI, T. 2012. MiR-96 downregulates REV1 and RAD51 to promote cellular sensitivity to cisplatin and PARP inhibition. *Cancer Res*, 72, 4037-46.
- WATSON, J. D. 1972. Origin of concatemeric T7 DNA. *Nat New Biol*, 239, 197-201.
- WENG, N. P., GRANGER, L. & HODES, R. J. 1997. Telomere lengthening and telomerase activation during human B cell differentiation. *Proc Natl Acad Sci U S A*, 94, 10827-32.
- WETERINGS, E. & CHEN, D. J. 2008. The endless tale of non-homologous end-joining. *Cell Res*, 18, 114-24.
- WHO 2014. CHRONIC LYMPHOCYTIC LEUKEMIA. Union for International Cancer Control. Review of Cancer Medicines on the WHO List of Essential Medicines.
- WILLIAMS, J., HEPPEL, N. H., BRITT-COMPTON, B., GRIMSTEAD, J. W., JONES, R. E., TAURO, S., BOWEN, D. T., KNAPPER, S., GROVES, M., HILLS, R. K., PEPPER, C., BAIRD, D. M. & FEGAN, C. 2017. Telomere length is an independent prognostic marker in MDS but not in de novo AML. *Br J Haematol*, 178, 240-249.
- WILLIAMSON, J. R., RAGHURAMAN, M. K. & CECH, T. R. 1989. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, 59, 871-80.

- WISE, J. L., CROUT, R. J., MCNEIL, D. W., WEYANT, R. J., MARAZITA, M. L. & WENGER, S. L. 2009. Human telomere length correlates to the size of the associated chromosome arm. *PLoS One*, 4, e6013.
- WONG, P., IWASAKI, M., SOMERVILLE, T. C., SO, C. W. & CLEARY, M. L. 2007. Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. *Genes Dev*, 21, 2762-74.
- WRIGHT, W. E., PEREIRA-SMITH, O. M. & SHAY, J. W. 1989. Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. *Mol Cell Biol*, 9, 3088-92.
- WU, P., TAKAI, H. & DE LANGE, T. 2012. Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell*, 150, 39-52.
- WYATT, D. W., FENG, W., CONLIN, M. P., YOUSEFZADEH, M. J., ROBERTS, S. A., MIECZKOWSKI, P., WOOD, R. D., GUPTA, G. P. & RAMSDEN, D. A. 2016. Essential Roles for Polymerase theta-Mediated End Joining in the Repair of Chromosome Breaks. *Mol Cell*, 63, 662-73.
- XIE, A., KWOK, A. & SCULLY, R. 2009. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat Struct Mol Biol*, 16, 814-8.
- YAKTAPOUR, N., UBELHART, R., SCHULER, J., AUMANN, K., DIERKS, C., BURGER, M., PFEIFER, D., JUMAA, H., VEELKEN, H., BRUMMER, T. & ZIRLIK, K. 2013. Insulin-like growth factor-1 receptor (IGF1R) as a novel target in chronic lymphocytic leukemia. *Blood*, 122, 1621-33.
- YANG, L., SHI, T., LIU, F., REN, C., WANG, Z., LI, Y., TU, X., YANG, G. & CHENG, X. 2015. REV3L, a promising target in regulating the chemosensitivity of cervical cancer cells. *PLoS One*, 10, e0120334.
- YANG, Q., ZHENG, Y. L. & HARRIS, C. C. 2005. POT1 and TRF2 cooperate to maintain telomeric integrity. *Mol Cell Biol*, 25, 1070-80.
- YATSENKO, S. A., HIXSON, P., RONEY, E. K., SCOTT, D. A., SCHAAF, C. P., NG, Y. T., PALMER, R., FISHER, R. B., PATEL, A., CHEUNG, S. W. & LUPSKI, J. R. 2012. Human subtelomeric copy number gains suggest a DNA replication mechanism for formation: beyond breakage-fusion-bridge for telomere stabilization. *Hum Genet*, 131, 1895-910.
- YE, J. Z. & DE LANGE, T. 2004. TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex. *Nat Genet*, 36, 618-23.
- YEAGER, T. R., NEUMANN, A. A., ENGLEZOU, A., HUSCHTSCHA, L. I., NOBLE, J. R. & REDDEL, R. R. 1999. Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body. *Cancer Res*, 59, 4175-9.
- YI, S., LI, Z., ZOU, D., AN, G., CUI, R., ZHONG, S., LI, H., XIONG, W., LI, C., CHEN, W., LIU, W., LV, R., YU, Z., WANG, H., XU, Y., ZHOU, K., RU, K., WANG, J., CHENG, T. & QIU, L. 2017. Intratumoral genetic heterogeneity and number of cytogenetic aberrations provide additional prognostic significance in chronic lymphocytic leukemia. *Genet Med*, 19, 182-191.
- YU, H. C., COUGHLIN, C. R., GEIGER, E. A., SALVADOR, B. J., ELIAS, E. R., CAVANAUGH, J. L., CHATFIELD, K. C., MIYAMOTO, S. D. & SHAIKH, T. H. 2016. Discovery of a potentially deleterious variant in TMEM87B in a patient with a hemizygous 2q13 microdeletion suggests a recessive condition characterized by congenital heart disease and restrictive cardiomyopathy. *Cold Spring Harb Mol Case Stud*, 2, a000844.
- YU, H. E., HAWASH, K., PICKER, J., STOLER, J., URION, D., WU, B. L. & SHEN, Y. 2012. A recurrent 1.71 Mb genomic imbalance at 2q13 increases the risk of developmental delay and dysmorphism. *Clin Genet*, 81, 257-64.

- ZHANG, F., KHAJAVI, M., CONNOLLY, A. M., TOWNE, C. F., BATISH, S. D. & LUPSKI, J. R. 2009. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet*, 41, 849-53.
- ZHANG, H., ZHU, L., HE, H., ZHU, S., ZHANG, W., LIU, X., ZHAO, X., GAO, C., MEI, M., BAO, S. & ZHENG, H. 2014. NF-kappa B mediated up-regulation of CCCTC-binding factor in pediatric acute lymphoblastic leukemia. *Mol Cancer*, 13, 5.
- ZHANG, J., VLASEVSKA, S., WELLS, V. A., NATARAJ, S., HOLMES, A. B., DUVAL, R., MEYER, S. N., MO, T., BASSO, K., BRINDLE, P. K., HUSSEIN, S., DALLA-FAVERA, R. & PASQUALUCCI, L. 2017. The CREBBP Acetyltransferase Is a Haploinsufficient Tumor Suppressor in B-cell Lymphoma. *Cancer Discov*, 7, 322-337.
- ZHANG, Q., KIM, N. K. & FEIGON, J. 2011a. Architecture of human telomerase RNA. *Proc Natl Acad Sci U S A*, 108, 20325-32.
- ZHANG, W., EDWARDS, A., FAN, W., DEININGER, P. & ZHANG, K. 2011b. Alu distribution and mutation types of cancer genes. *BMC Genomics*, 12, 157.
- ZHANG, Y. & JASIN, M. 2011. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nat Struct Mol Biol*, 18, 80-4.
- ZHONG, F. L., BATISTA, L. F., FREUND, A., PECH, M. F., VENTEICHER, A. S. & ARTANDI, S. E. 2012. TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell*, 150, 481-94.
- ZHU, J., ZHAO, Y. & WANG, S. 2010. Chromatin and epigenetic regulation of the telomerase reverse transcriptase gene. *Protein Cell*, 1, 22-32.
- ZHU, Y., MCAVOY, S., KUHN, R. & SMITH, D. I. 2006. RORA, a large common fragile site gene, is involved in cellular stress response. *Oncogene*, 25, 2901-8.
- ZOU, L. & ELLEDGE, S. J. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science*, 300, 1542-8.
- ZSCHENKER, O., KULKARNI, A., MILLER, D., REYNOLDS, G. E., GRANGER-LOCATELLI, M., POTTIER, G., SABATIER, L. & MURNANE, J. P. 2009. Increased sensitivity of subtelomeric regions to DNA double-strand breaks in a human cancer cell line. *DNA Repair (Amst)*, 8, 886-900.